Statistics in Psychology

Block-1 Introduction to Statistics [4]

Unit-1 Parametric and Non-parametric Statistics

Unit-2 Descriptive and Inferential Statistics

Unit-3 Type I and Type II Errors

Unit-4 Setting Up the Levels of Significance

Block-2 Correlation and Regression [4]

Unit-1 Product Moment Coefficient of Correlation

Unit-2 Other Types of Correlations (phi-coefficient)

Unit-3 Partial and Multiple Correlations

Unit-4 Bivariate and Multiple Regression

Block-3 Normal Distribution [4]

Unit-1 Characteristics of Normal Distribution

Unit-2 Significance of Mean Differences, Standard Error of the Mean

Unit-3 One Way Analysis of Variance

Unit-4 Two Way Analysis of Variance

Block-4 Non-Parametric Statistics [4]

Unit-1 Rationale for Non-parametric Statistics

Unit-2 Mann Whitney 'U' Test for Two Sample Test

Unit-3 Kruskal Wallis Analysis of Variance

Unit-4 Chi-Square and Kendall Rank Correlation

SHOUT
SHARE
IT

UNIT 1 PARAMETRIC AND NON-PARAMETRIC STATISTICS

Structure

1.0 Introduction

1.1 Objectives

1.2 Definition of Parametric and Non-parametric Statistics

1.3 Assumptions of Parametric and Non-parametric Statistics

- 1.3.1 Assumptions of Parametric Statistics
- 1.3.2 Assumptions of Non-parametric Statistics

1.4 Advantages of Non-parametric Statistics

1.5 Disadvantages of Non-parametric Statistical Tests

1.6 Parametric Statistical Tests for Different Samples

1.7 Parametric Statistical Measures for Calculating the Difference Between Means

- 1.7.1 Significance of Difference Between the Means of Two Independent Large and Small Samples
- 1.7.2 Significance of the Difference Between the Means of Two Dependent Samples
- 1.7.3 Significance of the Difference Between the Means of Three or More Samples

1.8 Parametric Statistics Measures Related to Pearson's 'r'

- 1.8.1 Non-parametric Tests Used for Inference

1.9 Some Non-parametric Tests for Related Samples

1.10 Let Us Sum Up

1.11 Unit End Questions

1.12 Glossary

1.13 Suggested Readings

1.0 INTRODUCTION

In this unit you will be able to know the various aspects of parametric and non-parametric statistics. A parametric statistical test specifies certain conditions such as the data should be normally distributed etc. The non-parametric statistics does not require the conditions of parametric stats. In fact non-parametric tests are known as distribution free tests.

In this unit we will study the nature of quantitative data and various descriptive statistical measures which are used in the analysis of such data. These include measures of central tendency, variability, relative position and relationships of normal probability curve etc. will be explained.

The computed values of various statistics are used to describe the properties of particular samples. In this unit we shall discuss inferential or sampling statistics, which are useful to a researcher in making generalisations of inferences about the populations from the observations of the characteristics of samples.

For making inferences about various population values (parameters), we generally

make use of parametric and non-parametric tests. The concept and assumptions of parametric tests will be explained to you in this section along with the inference regarding the means and correlations of large and small samples, and significance of the difference between the means and correlations in large and small independent samples.

The assumptions and applications of analysis of variance and co-variance for testing the significance of the difference between the means of three or more samples will also be discussed.

In the use of parametric tests for making statistical inferences, we need to take into account certain assumptions about the nature of the population distribution, and also the type of the measurement scale used to quantify the data. In this unit you will learn about another category of tests which do not make stringent assumptions about the nature of the population distribution. This category of test is called distribution free or non-parametric tests. The use and application of several non-parametric tests involving unrelated and related samples will be explained in this unit. These would include chi-square test, median test, Man-Whitney U test, sign test and Wilcoxon-matched pairs signed-ranks test.

1.1 OBJECTIVES

After reading this unit, you will be able to:

- define the terms parametric and non-parametric statistics;
- differentiate between parametric and non-parametric statistics;
- describe the nature and meaning of parametric and non-parametric statistics;
- delineate the assumptions of parametric and non-parametric statistics; and
- list the advantages and disadvantages of parametric and non-parametric statistics.

1.2 DEFINITION OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

Statistics is an Independent branch and its use is highly prevalent in all the fields of knowledge. Many methods and techniques are used in statistics. These have been grouped under parametric and and non-parametric statistics. Statistical tests which are not based on a *normal distribution* of data or on any other assumption are also known as distribution-free tests and the data are generally ranked or grouped. Examples include the *chi-square test* and *Spearman's rank correlation coefficient*.

The first meaning of *non-parametric* covers techniques that do not rely on data belonging to any particular distribution. These include, among others:

1) Distribution free methods: This means that there are no assumptions that the data have been drawn from a normally distributed population. This consists of non-parametric *statistical models*, *inference* and *statistical tests*.

2) Non-parametric statistics: In this the statistics is based on the *ranks* of observations and do not depend on any distribution of the population.

3) No assumption of a structure of a model: In non-parametric statistics, the techniques do not assume that the *structure* of a model is fixed. In this, the

individual variables *are* typically assumed to belong to parametric distributions, and assumptions about the types of connections among variables are also made. These techniques include, among others:

a) Non-parametric regression

b) Non-parametric hierarchical Bayesian models.

In non-parametric regression, the structure of the relationship is treated non-parametrically.

In regard to the Bayesian models, these are based on the *Dirichlet process*, which allows the number of *latent variables* to grow as necessary to fit the data. In this the individual variables however follow parametric distributions and even the process controlling the rate of growth of latent variables follows a parametric distribution.

4) The assumptions of a *classical* or *standard tests* are not applied to non-parametric tests.

Parametric tests

Parametric tests normally involve data expressed in absolute numbers or values rather than ranks; an example is the *Student's t-test*.

The parametric statistical test operates under certain conditions. Since these conditions are not ordinarily tested, they are assumed to hold valid. The meaningfulness of the results of a parametric test depends on the validity of the assumption. Proper interpretation of parametric test based on normal distribution also assumes that the scene being analysed results from measurement in at least an interval scale.

Let us try to understand the term population. Population refers to the entire group of people which a researcher intends to understand in regard to a phenomenon. The study is generally conducted on a sample of the said population and the obtained results are then applied to the larger population from which the sample was selected.

Tests like t, z, and F are called parametrical statistical tests.

T-tests: A T-test is used to determine if the scores of two groups differ on a single variable.

A t-test is designed to test for the differences in mean scores. For instance, you could use t-test to determine whether writing ability differs among students in two classrooms.

It may be mentioned here that the parametric tests, namely, t-test and F-test, are considered to be quite robust and are appropriate even when some assumptions are not met.

Parametric tests are useful as these tests are most powerful for testing the significance or trustworthiness of the computed sample statistics. However, their use is based upon certain assumptions. These assumptions are based on the nature of the population distribution and on the way the type of scale is used to quantify the data measures.

Let us try to understand what is a scale and its types. There are four types of scales used in measurement viz., nominal scale, ordinal scale, interval scale, and ratio scale.

1) Nominal scale deals with nominal data or classified data such as for example the population divided into males and females. There is no ordering of the data in that it has no meaning when we say male > female. These data are also given

arbitrary labels such as m / f and 1 //0 . These are also called as categorical scale , that is these are scales with values that are in terms of categories (i.e. they are names rather than numbers).

2) Ordinal scale deals with interval data. These are in certain order but the differences between values are not important. For example, degree of satisfaction ranging in a 5 point scale of 1 to 5, with 1 indicating least satisfaction and 5 indicating high satisfaction.

3) Interval scale deals with ordered data with interval. This is a constant scale but has no natural zero. Differences do make sense . Example of this kind of data includes for instance temperature in Centigrade or Fahrenheit. The dates in a calendar. Interval scale possesses two out of three important requirements of a good measurement scale, that is, magnitude and equal intervals but lacks the real or absolute zero point.

4) Ratio scale deals with ordered, constant scale with a natural zero. Example of this type of data include for instance, height, weight, age, length etc.

The sample with small number of items are treated with non-parametric statistics because of the absence of normal distribution, e.g. if our sample size is 30 or less; ($N \leq 30$). It can be used even for nominal data along with the ordinal data.

A non-parametric statistical test is based on model that specifies only very general conditions and none regarding the specific form of the distribution from which the sample was drawn.

Certain assumptions are associated with most non-parametric statistical tests, namely that the observation are independent, perhaps that variable under study had underlying continuity, but these assumptions are fewer and weaker than those associated with parametric tests.

More over as we shall see, non-parametric procedures often test different hypotheses about population than do parametric procedures.

Finally, unlike parametric tests, there are non-parametric procedures that may be applied appropriately to data measured in an ordinal scale, or in a nominal scale or categorical scale.

Non-parametric statistics deals with small sample sizes.

Non-parametric statistics are assumption free meaning these are not bound by any assumptions.

Non-parametric statistics are user friendly compared with parametric statistics and economical in time.

We have learnt that parametric tests are generally quite robust and are useful even when some of their mathematical assumptions are violated. However, these tests are used only with the data based upon ratio or interval measurements.

In case of counted or ranked data, we make use of non-parametric tests. It is argued that non-parametric tests have greater merit because their validity is not based upon assumptions about the nature of the population distribution, assumptions that are so frequently ignored or violated by researchers using parametric tests. It may be noted that non-parametric tests are less precise and have less power than the parametric tests.

1.3 ASSUMPTIONS OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

1.3.1 Assumptions of Parametric Statistics

Parametric tests like, 't and f' tests may be used for analysing the data which satisfy the following conditions :

The population from which the sample have been drawn should be normally distributed.

Normal Distributions refer to Frequency distribution following a normal curve, which is infinite at both the ends.

The variables involved must have been measured interval or ratio scale.

Variable and its types: characteristic that can have different values.

Types of Variables

Dependent Variable: Variable considered to be an effect; usually a measured variable.
Independent Variable: Variable considered being a cause.

The observation must be independent. The inclusion or exclusion of any case in the sample should not unduly affect the results of study.

These populations must have the same variance or, in special cases, must have a known ratio of variance. This we call homosedasticity.

The samples have equal or nearly equal variances. This condition is known as equality or homogeneity of variances and is particularly important to determine when the samples are small.

The observations are independent. The selection of one case in the sample is not dependent upon the selection of any other case.

1.3.2 Assumptions of Non-parametric Statistics

We face many situations where we can not meet the assumptions and conditions and thus cannot use parametric statistical procedures. In such situation we are bound to apply non-parametric statistics.

If our sample is in the form of nominal or ordinal scale and the distribution of sample is not normally distributed, and also the sample size is very small, it is always advisable to make use of the non-parametric tests for comparing samples and to make inferences or test the significance or trust worthiness of the computed statistics. In other words, the use of non-parametric tests is recommended in the following situations:

Where sample size is quite small. If the size of the sample is as small as N=5 or N=6, the only alternative is to make use of non-parametric tests.

When assumption like normality of the distribution of scores in the population are doubtful, we use non-parametric tests.

When the measurement of data is available either in the form of ordinal or nominal scales or when the data can be expressed in the form of ranks or in the shape of + signs or – signs and classification like "good-bad", etc., we use non-parametric statistics.

The nature of the population from which samples are drawn is not known to be normal.

The variables are expressed in nominal form.

The data are measures which are ranked or expressed in numerical scores which have the strength of ranks.

1.4 ADVANTAGES OF NON-PARAMETRIC STATISTICS

If the sample size is very small, there may be no alternative except to use a non-parametric statistical test.

Non-parametric tests typically make fewer assumptions about the data and may be relevant to a particular situation.

The hypothesis tested by the non-parametric test may be more appropriate for research investigation.

Non-parametric statistical tests are available to analyse data which are inherently in ranks as well as data whose seemingly numerical scores have the strength of ranks.

For example, in studying a variable such as anxiety, we may be able to state that subject A is more anxious than subject B without knowing at all exactly how much more anxious A is. Thus if the data are inherently in ranks, or even if they can be categorised only as plus or minus (more or less, better or worse), they can be treated by non-parametric methods.

Non-parametric methods are available to treat data which are simply classificatory and categorical, i.e., are measured in nominal scale.

Samples made up of observations from several different populations at times cannot be handled by Parametric tests.

Non-parametric statistical tests typically are much easier to learn and to apply than are parametric tests. In addition, their interpretation often is more direct than the interpretation of parametric tests.

1.5 DISADVANTAGES OF NON-PARAMETRIC STATISTICAL TESTS

If all the assumptions of a parametric statistical model are in fact met in the data and the research hypothesis could be tested with a parametric test, then non-parametric statistical tests are wasteful.

The degree of wastefulness is expressed by the power-efficiency of the non-parametric test. It will be remembered that, if a non-parametric statistical test has power-efficiency of say, 90 percent, this means that when all conditions of parametric statistical test are satisfied the appropriate parametric test would be just as effective with a sample which is 10 percent smaller than that used in non-parametric analysis.

Another objection to non-parametric statistical test has to do with convenience. Tables necessary to implement non-parametric tests are scattered widely and appear in different formats (The same is true of many parametric tests too).

1.6 PARAMETRIC STATISTICAL TESTS FOR DIFFERENT SAMPLES

Suppose we wish to measure teaching aptitude of M.A. Psychology Students(LARGE SAMPLE) by using a verbal aptitude teaching test.

It is not possible and convenient to measure the teaching aptitude of all the enrolled M.A. Psychology Students trainees and hence we must usually be satisfied with a sample drawn from this population.

However, this sample should be as large and as randomly drawn as possible so as to represent adequately all the M.A. Psychology Students of IGNOU.

If we select a large number of random samples of 100 trainees each from the population of all trainees, the mean values of teaching aptitude scores for all samples would not be identical.

A few would be relatively high, a few relatively low, but most of them would tend to cluster around the population mean.

The sample means due to 'sampling error' will not vary from sample to sample but will also usually deviate from the population mean. Each of these sample means can be treated as a single observation and these means can be put in a frequency distribution which is known as sampling distribution of the means.

An important principle, known as the 'Central Limit Theorem', describes the characteristics of sample means. According to this theorem, if a large number of equal-sized samples, greater than 30 in size, are selected at random from an infinite population:

The means of the samples will be normally distributed.

The average value of the sample means will be the same as the mean of the population.

The distribution of sample means will have its own standard deviation.

This standard deviation is known as the 'standard error of the mean' which is denoted as SE_M or $ó_M$.

It gives us a clue as to how far such sample means may be expected to deviate from the population mean.

The standard error of a mean tells us how large the errors are in any particular sampling situation.

The formula for the standard error of the mean in a large sample is:

$$SE_M \text{ or } ó_M = ó / \sqrt{N}$$

Where

ó = the standard deviation of the population

N = the size of the sample

In case of **small samples**, the sampling distribution of means is not normal. It was in about 1815 when William Seely Gosset developed the concept of small sample size. He found that the distribution curves of small sample means were some what different from the normal curve. This distribution was named as t-distribution. When the size of the sample is small, the t-distribution lies under the normal curve.

1.7 PARAMETRIC STATISTICAL MEASURES FOR CALCULATING DIFFERENCE BETWEEN MEANS

In some research situations we require the use of a statistical technique to determine whether a true difference exists between the population parameters of two samples. The parameters may be means, standard deviations, correlations etc. For example, suppose we wish to determine whether the population of male M.A. Psychology Students enrolled with IGNOU differs from their female counterparts in their attitude towards teaching… In this case we would first draw samples of male and female M.A. Psychology Students. Next, we would administer an attitude scale measuring attitude towards teaching on the selected samples, compute the means of the two samples, and find the difference between them. Let the mean of the male sample be 55 and that of the females 59. Then it has to be ascertained if the difference of 4 between the sample means is large enough to be taken as real and not due only to sampling error or chance.

In order to test the significance of the obtained difference of 4, we need to first find out the standard error of the difference of the two means because it is reasonable to expect that the difference between two means will be subject to sampling errors. Then from the difference between the sample means and its standard error we can determine whether a difference probably exists between the population means.

In the following sections we will discuss the procedure of testing the significance of the difference between the means and correlations of the samples.

1.7.1 Significance of the Difference between the Means of Two Independent Large and Small Samples

Means are said to be independent or uncorrelated when computed from samples drawn at random from totally different and unrelated groups.

Large Samples

You have leant that the frequency distribution of large sample means, drawn from the same population, fall into a normal distribution around the population mean (M_{pop}) as their measure of central tendency. It is reasonable to expect that the frequency distribution of the difference between the means computed from the samples drawn from two different populations will also tend to be normal with a mean of zero and standard deviation which is called the standard error of the difference of means.

The standard error is denoted by $ó_{dm}$ which is estimated from the standard errors of the two sample means, $ó_{m1}$ and $ó_{m2}$. The formula is:

$ó_{dM} = (ó_{M1}2 + ó_{M1}2)^{\text{Under root}}$

in which

$ó_{M1}$ = SE of the mean of the first sample

$ó_{Mw}$ = SE of the mean of the second sample

N_1 = Number of cases in first sample

N_2 = Number of cases in second sample

1.7.2 Significance of the Differences between the Means of Two Dependent Samples

Means are said to be dependent or correlated when obtained from the scores of the same test administered to the same sample upon two occasions, or when the same test is administered to equivalent samples in which the members of the group have been matched person for person, by one or more attributes.

$T = M_1 + M_2 / \acute{o}M^2_1 + - 2\ r_{12}\ \acute{o}M_1\ \acute{o}M_2$

in which

M_1 and M_2 = Means of the scores of the initial and final testing.

$\acute{o}M_1$ = Standard error of the initial test mean.

$\acute{o}M_2$ = Standard error of the final test mean.

r_{12} = Correlation between the scores on initial and final testing.

1.7.3 Significance of the Difference between the Means of Three or More Samples

We compute CR and t-values to determine whether there is any significant difference between the means of two random samples. Suppose we have N(N>2) random samples and we want to determine whether there are any significant differences among their means. For this we have to compute F value that is Analysis of Variance.

Analysis of variance has the following basic assumptions underlying it which should be fulfilled in the use of this technique.

The population distribution should be normal. This assumption, however, is not especially important.

Eden and Yates showed that even with a population departing considerably from normality, the effectiveness of the normal distribution still held.

All the groups of certain criterion or of the combination of more than one criterion should be randomly chosen from the sub-population having the same criterion or having the same combination of more than one criterion.

For instance, if we wish to select two groups in a population of M.A. Psychology Student trainees enrolled with IGNOU, one of males and the other of females, we must choose randomly from the respective sub populations. The assumption of randomness is the key stone of the analysis of variance technique. There is no substitute for randomization.

The sub-groups under investigation should have the same variability. This assumption is tested by applying F_{max} test.

F_{max} =Largest Variance / Smallest Variance

In analysis of variance, we have usually three or more groups i.e. there will be three or more variances.

Unless the computed value of F_{max} equals or exceeds the appropriate F critical value at .05 level in the Table N of the Appendix, (Statistics book) it is assumed that the variances are homogeneous and the difference is not significant.

1.8 PARAMETRIC STATISTICS MEASURES RELATED TO PEARSON'S 'r'

The mathematical basis for standard error of a Pearson's co-efficient of correlation 'r' is rather complicated because of the difficulty in its nature of sampling distribution.

The sampling distribution of r is not normal except when population r is near zero and size of the sample is large (N=30 or greater).

When r is high (0.80 or more) and N is small, the sampling distribution of r is skewed. It is also true when r is low (0.20 or less).

In view of this, a sound method for making the inference regarding Pearson's r, especially when its magnitude is very high or very low, is to convert r into Fisher's Z coefficient using conversion table provided in the Appendix (Statistics book) and find the standard error (SE) of Z.

The sampling distribution of Z co-efficient is normal regardless of the size of sample N and the size of the population r. Furthermore, the SE of Z depends only upon the size of sample N.

The formula for standard error of Z ($ó_z$) is:

$$SE_z = 1/\sqrt{N-3}$$

The method of determining the standard error of the difference between Pearson's co-efficient of correlation of two samples is first to convert the r's into Fisher's Z co-efficient and then to determine the significance of the difference between the two Z's.

When we have two correlations between the same two variables, X and Y, computed from two totally different and unmatched samples, the standard error of a difference between two corresponding Z's is computed by the formula:

$$SE_{dz} = ó_{z1-z2} = \sqrt{(1/N_{1-}\ 3 + 1/\ N_{2-}\ 3)}$$

in which

N_1 and N_2 = sizes of the two samples

The significance of the difference between the two Z's is tested with the following formula:

$$CR = Z_1 - Z_2\ /\ SE_{DZ}$$

1.8.1 Non-parametric Tests Used for Inference

The most frequently used non-parametric tests for drawing statistical inferences in case of unrelated or independent samples are:

1) Chi square test;

2) Median test; and

3) Mann-Whitney 'U' test.

The use and application of these tests are discussed below:

The Chi Square (X^2) Test

The chi square test is applied only to discrete data. The data that are counted rather

than measured. It is a test of independence and is used to estimate the likelihood that some factor other than chance accounts for the observed relationship.

The Chi square (X^2) is not a measure of the degree of relationship between the variables under study.

The Chi square test merely evaluates the probability that the observed relationship results from chance. The basic assumption, as in case of other statistical significance, is that the sample observations have been randomly selected.

The formula for chi-square (X^2) is:

$$(X^2) = \sum[(fo-fe)2 / fe]$$

In which

Fo = frequency of occurrence of observed or experimentally determined facts.

Fe = expected frequency of occurrence.

The Median Test

The median test is used for testing whether two independent samples differ in central tendencies. It gives information as to whether it is likely that two independent samples have been drawn from populations with the same median. It is particularly useful when even the measurements for the two samples are expressed in an ordinal scale. In using the median test, we first calculate the combined median for all measures (scores) in both samples. Then both sets of scores at the combined median are dichotomized and the data are set in a 2 x 2 table with two rows one containing below median and the other row containing above median. On the column side we have two columns, one containing the sample 1 and the other column containing sample 2.

The Mann-Whitney U Test

The Mann-Whitney U test is more useful than the Median test. It is one of the most useful alternative to the parametric t test when the parametric assumptions cannot be met and when the measurements are expressed in ordinal scale values.

1.9 SOME NON-PARAMETRIC TESTS FOR RELATED SAMPLES

Various tests are used in drawing statistical inferences in case of related samples. In this section we shall confine our discussion to the use of Sign Test and Wilcoxon Matched-Pairs Signed-Ranks Test Only

The Sign Test

The sign test is the simplest test of significance in the category of non-parametric tests. It makes use of plus and minus signs rather than quantitative measures as its data. It is particularly useful in situations in which quantitative measurement is impossible or inconvenient, but on the basis of superior or inferior performance it is possible to rank with respect to each other, the two members of each pair.

The sign test is used either in the case of single sample from which observations are obtained under two experimental conditions. The researcher wants to establish that the two conditions are different.

The use of this test does not make any assumption about the form of the distribution of differences. The only assumption underlying this test is that the variable under investigation has a continuous distribution.

The Wilcoxon Matched Pairs Signed Ranks Test

The Wilcoxon matched pairs signed ranks test is more powerful than the sign test because it tests not only direction but also the magnitude of differences within pairs of matched groups.

This test, like the sign test, deals with dependent groups made up of matched pairs of individuals and is not applicable to independent groups. The null hypothesis would assume that the direction and magnitude of pair difference would be about the same.

1.10 LET US SUM UP

Parametric and non-parametric tests are important for students especially researchers working in any field. Parametric tests include all methods of statistics when the sample size is large whereas in non-parametric test the sample size is small. There are some advantages and disadvantages of both the tests. In this unit we discussed the statistical inference based on parametric tests. It included the assumptions on which the use of parametric tests are based; inferences regarding means of large and small samples; significance of the difference between the means of two large and small independent samples; significance of the difference between means of the two dependent samples; significance of the difference between means of three or more samples; significance of Pearson's coefficients of correlation; and significance of the difference between Pearson's coefficients of correlation of two independent samples. F test is used for testing the significance between the means of three or more samples. It involves the use of analysis of variance or analysis of co-variance. For testing the significance of Pearson's r, we make use of Fisher's Z transformation or t-test.

1.11 UNIT END QUESTIONS

1) Define parametric statistics.

2) Discuss non-parametric statistics?

3) Write various assumptions of parametric statistics?

4) What are the advantages of non-parametric statistics?

5) Differentiate between parametric and non-parametric statistics?

6) List the assumptions on which the use of Parametric Tests is base.

7) Describe the characteristics of Central Limit Theorem.

8) Define the standard error of mean.

1.12 GLOSSARY

Statistics : Measurement which are associated with sample

Parameters : Measurements which are associated with population

Assumptions	:	Prerequisite conditions
Population	:	Larger group of people to which inferences are made.
Sample	:	Small proportion of the population which we assert representing population.
Normal Curve	:	Bell shaped frequency distribution that is symmetrical and unimodel.
Distribution free tests	:	Hypothesis – testing procedure making non assumptions about population parameters.
Categorical Scale	:	Variable with values that are categories that is, they are name rather than numbers.
Test	:	Test is a tool to measure observable behaviour
Homosedasity	:	Populations must have some variance or in special cases must have a known ratio of variance.

1.13 SUGGESTED READINGS

Asthana H.S, and Bhushan. B. (2007) *Statistics for Social Sciences* (with SPSS Applications).

B.L. Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Guilford, J.P. (1965); *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill Book Company.

Siegel, S. (1956): *Non-parametric Statistics for Behavioural Sciences*. Tokyo: McGraw Hill Hoga Kunsa Ltd.

Sidney Siegel, & N. John Castetellan, Jr. (1958) *Non-parametric Statistics for the Behavioural Science*. McGraw Hill Books company, New Delhi

UNIT 2 DESCRIPTIVE AND INFERENTIAL STATISTICS

Structure

2.0 Introduction

2.1 Objectives

2.2 Meaning of Descriptive Statistics

2.3 Organisation of Data

- 2.3.1 Classification
  - 2.3.1.1 Frequency Distribution can be with Ungrouped Data and Grouped Data
  - 2.3.1.2 Types of Frequency Distribution
- 2.3.2 Tabulation
- 2.3.3 Graphical Presentation of Data
  - 2.3.3.1 Cumulative Frequency Curve or Ogive
- 2.3.4 Diagrammatic Presentation of Data

2.4 Summarisation of Data

- 2.4.1 Measures of Central Tendency
- 2.4.2 Measures of Dispersion
- 2.4.3 Skewness and Kurtosis
- 2.4.4 Advantages and Disadvantages of Descriptive Statistics

2.5 Meaning of Inferential Statistics

- 2.5.1 Estimation
- 2.5.2 Point Estimation
- 2.5.3 Interval Estimation

2.6 Hypothesis Testing

- 2.6.1 Statement of Hypothesis
- 2.6.2 Level of Significance
- 2.6.3 One Tail and Two Tail Test

2.7 Errors in Hypothesis Testing

- 2.7.1 Type I Error
- 2.7.2 Type II Error
- 2.7.3 Power of a Test

2.8 General Procedure for Testing A Hypothesis

2.9 Let Us Sum Up

2.10 Unit End Questions

2.11 Glossary

2.12 Suggested Readings

2.0 INTRODUCTION

In this unit we will be dealing with descriptive and inferential statistics. First we start with defining descriptive statistics and indicate how to organise the data , classify, tabulate etc. This unit also presents as to how the data should be presented graphically. Once the data is collected the same has to be made meaningful which can be done through averaging the data or working out the variances in the data etc. Then we deal with the advantages and disadvantages of descriptive statistics. This is followed

by defining what is inferential statistics and delineating its meaning. In this unit the student will also gain knowledge regarding point and interval estimation so as to validate the results. We also learn in this unit about hypothesis testing, how it is done and the methods thereof. We also deal with different types of errors in hypothesis testing including sampling error etc.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- define the nature and meaning of descriptive statistics;
- describe the methods of organising and condensing raw data;
- explain concept and meaning of different measures of central tendency;
- analyse the meaning of different measures of dispersion;
- define inferential statistics;
- explain the concept of estimation;
- distinguish between point estimation and interval estimation; and
- explain the different concepts involved in hypothesis testing.

2.2 MEANING OF DESCRIPTIVE STATISTICS

The word statistics has different meaning to different persons. For some, it is a one-number description of a set of data. Some consider statistics in terms of numbers used as measurements or counts. Mathematicians use statistics to describe data in one word. It is a summary of an event for them. Number , n, is the statistic describing how big the set of numbers is, how many pieces of data are in the set.

Also, knowledge of statistics is applicable in day to day life in different ways. Statistics is used by people to take decision about the problems on the basis of different types of information available to them. However, in behavioural sciences the word 'statistics' means something different, that is its prime function is to draw statistical inference about population on the basis of available quantitative and qualitative information.

The word statistics can be defined in two different ways. In singular sense 'Statistics' refers to what is called statistical methods. When 'Statistics' is used in plural sense it refers to 'data'.

In this unit we will use the term 'statistics' in singular sense. In this context, it is described as a branch of science which deals with the collection of data, their classification, analysis and interpretations of statistical data.

The science of statistics may be broadly studied under two headings:

i) Descriptive Statistics, and (ii) Inferential Statistics

i) Descriptive Statistics: Most of the observations in this universe are subject to variability, especially observations related to human behaviour. It is a well known fact that attitude, intelligence and personality differ from individual to individual. In order to make a sensible definition of the group or to identify the group with reference to their observations/ scores, it is necessary to express them in a precise manner. For this purpose observations need to be expressed as a single estimate which summarises the observations.

Descriptive statistics is a branch of statistics, which deals with descriptions of obtained data. On the basis of these descriptions a particular group of population is defined for corresponding characteristics. The descriptive statistics include classification, tabulation, diagrammatic and graphical presentation of data, measures of central tendency and variability. These measures enable the researchers to know about the tendency of data or the scores, which further enhance the ease in description of the phenomena. Such single estimate of the series of data which summarises the distribution are known as parameters of the distribution. These parameters define the distribution completely.

Bascially descriptive statistics involves two operations:

(i) Organisation of data, and (ii) Summarination of data

2.3 ORGANISATION OF DATA

There are four major statistical techniques for organising the data. These are:

i) Classification

ii) Tabulation

iii) Graphical Presentation, and

iv) Diagrammatical Presentation

2.3.1 Classification

The arrangement of data in groups according to similarities is known as classification. A classification is a summary of the frequency of individual scores or ranges of scores for a variable. In the simplest form of a distribution, we will have such value of variable as well as the number of persons who have had each value.

Once data are collected, it should be arranged in a format from which they would be able to draw some conclusions. Thus by classifying data, the investigators move a step ahead in regard to making a decision.

A much clear picture of the information of score emerges when the raw data are organised as a frequency distribution. Frequency distribution shows the number of cases following within a given class interval or range of scores. A frequency distribution is a table that shows each score as obtained by a group of individuals and how frequently each score occurred.

2.3.1.1 Frequency Distribution can be with Ungrouped Data and Grouped Data

i) An ungrouped frequency distribution may be constructed by listing all score values either from highest to lowest or lowest to highest and placing a tally mark (/) besides each scores every times it occurs. The frequency of occurrence of each score is denoted by ‘f’.

ii) Grouped frequency distribution: If there is a wide range of score value in the data, then it is difficult to get a clear picture of such series of data. In this case grouped frequency distribution should be constructed to have a clear picture of the data. A group frequency distribution is a table that organises data into classes.

It shows the number of observations from the data set that fall into each of the class.

Construction of frequency distribution

To prepare a frequency distribution it is essential to determine the following:

1) The range of the given data =, the difference between the highest and lowest scores.

2) The number of class intervals = There is no hard and fast rules regarding the number of classes into which data should be grouped. If there are very few scores it is useless to have a large number of class-intervals. Ordinarily, the number of classes should be between 5 to 30.

3) Limits of each class interval = Another factor used in determining the number of classes is the size/ width or range of the class which is known as 'class interval' and is denoted by 'i'.

Class interval should be of uniform width resulting in the same-size classes of frequency distribution. The width of the class should be a whole number and conveniently divisible by 2, 3, 5, 10, or 20.

There are three methods for describing the class limits for distribution:

(i) Exclusive method, (ii) Inclusive method and (iii) True or actual class method.

i) **Exclusive method**

In this method of class formation, the classes are so formed that the upper limit of one class become the lower limit of the next class. In this classification, it is presumed that score equal to the upper limit of the class is exclusive, i.e., a score of 40 will be included in the class of 40 to 50 and not in a class of 30 to 40 (30-40, 40-50, 50-60)

ii) **Inclusive method**

In this method the classes are so formed that the upper limit of one class does not become the lower limit of the next class. This classification includes scores, which are equal to the upper limit of the class. Inclusive method is preferred when measurements are given in whole numbers. (30-39, 40-49, 50-59)

iii) **True or Actual class method**

Mathematically, a score is an internal when it extends from 0.5 units below to 0.5 units above the face value of the score on a continuum. These class limits are known as true or actual class limits. (29.5 to 39.5, 39.5 to 49.5) etc.

2.3.1.2 Types of Frequency Distribution

There are various ways to arrange frequencies of a data array based on the requirement of the statistical analysis or the study. A couple of them are discussed below:

i) Relative frequency distribution: A relative frequency distribution is a distribution that indicates the proportion of the total number of cases observed at each score value or internal of score values.

ii) Cumulative frequency distribution: Sometimes investigator may be interested to know the number of observations less than a particular value. This is possible by computing the cumulative frequency. A cumulative frequency corresponding

to a class-interval is the sum of frequencies for that class and of all classes prior to that class.

iii) Cumulative relative frequency distribution: A cumulative relative frequency distribution is one in which the entry of any score of class interval expresses that score's cumulative frequency as a proportion of the total number of cases.

Self Assessment Questions

1) Complete the following statements

   i) Statistics in plural means .....

   ii) Statistics in singular means .....

   iii) Data collection is ..... step in statistics.

   iv) The last step in statistics is .....

2) Define following concepts

   1) Descriptive statistics

      .....

      .....

   2) Inferential statistics

      .....

      .....

   3) Exclusive method of classification

      .....

      .....

   4) Actual method of classification

      .....

      .....

   5) Frequency distribution

      .....

      .....

2.3.2 Tabulation

Frequency distribution can be either in the form of a table or it can be in the form of graph. Tabulation is the process of presenting the classified data in the form of a table. A tabular presentation of data becomes more intelligible and fit for further statistical analysis. A table is a systematic arrangement of classified data in row and columns with appropriate headings and sub-headings. The main components of a table are:

i) Table number: When there is more than one table in a particular analysis a table should be marked with a number for their reference and identification. The number should be written in the center at the top of the table.

ii) Title of the table: Every table should have an appropriate title, which describes the content of the table. The title should be clear, brief, and self-explanatory. Title of the table should be placed either centrally on the top of the table or just below or after the table number.

iii) Caption: Captions are brief and self-explanatory headings for columns. Captions may involve headings and sub-headings. The captions should be placed in the middle of the columns. For example, we can divide students of a class into males and females, rural and urban, high SES and Low SES etc.

iv) Stub: Stubs stand for brief and self-explanatory headings for rows.

v) Body of the table: This is the real table and contains numerical information or data in different cells. This arrangement of data remains according to the description of captions and stubs.

vi) Head note: This is written at the extreme right hand below the title and explains the unit of the measurements used in the body of the tables.

vii) Footnote: This is a qualifying statement which is to be written below the table explaining certain points related to the data which have not been covered in title, caption, and stubs.

viii) Source of data: The source from which data have been taken is to be mentioned at the end of the table.

TITLE

Stub Head Stub Entries	**Caption**			
	Column Head I		Column Head II	
	Sub Head	Sub Head	Sub Head	Sub Head
	MAIN BODY	OF	THE TABLE	
Total				

Footnote(s):

Source :

2.3.3 Graphical Presentation of Data

The purpose of preparing a frequency distribution is to provide a systematic way of " looking at" and understanding data. To extend this understanding, the information contained in a frequency distribution often is displayed in graphic and/or diagrammatic forms. In graphical presentation of frequency distribution, frequencies are plotted on a pictorial platform formed of horizontal and vertical lines known as graph.

A graph is created on two mutually perpendicular lines called the X and Y–axes on which appropriate scales are indicated. The horizontal line is called the abscissa and vertical the ordinate. Like different kinds of frequency distributions there are many kinds of graph too, which enhance the scientific understanding of the reader. The

commonly used graphs are Histogram, Frequency polygon, Frequency curve, Cumulative frequency curve. Here we will discuss some of the important types of graphical patterns used in statistics.

i) Histogram: It is one of the most popular methods for presenting continuous frequency distribution in a form of graph. In this type of distribution the upper limit of a class is the lower limit of the following class. The histogram consists of series of rectangles, with its width equal to the class interval of the variable on horizontal axis and the corresponding frequency on the vertical axis as its heights.

ii) Frequency polygon: Prepare an abscissa originating from 'O' and ending to 'X'. Again construct the ordinate starting from 'O' and ending at 'Y'. Now label the class-intervals on abscissa stating the exact limits or midpoints of the class-intervals. You can also add one extra limit keeping zero frequency on both side of the class-interval range.

The size of measurement of small squares on graph paper depends upon the number of classes to be plotted. Next step is to plot the frequencies on ordinate using the most comfortable measurement of small squares depending on the range of whole distribution.

To plot a frequency polygon you have to mark each frequency against its concerned class on the height of its respective ordinate. After putting all frequency marks a draw a line joining the points. This is the polygon.

iii) Frequency curve: A frequency curve is a smooth free hand curve drawn through frequency polygon. The objective of smoothing of the frequency polygon is to eliminate as far as possible the random or erratic fluctuations that are present in the data.

2.3.3.1 Cumulative Frequency Curve or Ogive

The graph of a cumulative frequency distribution is known as cumulative frequency curve or ogive. Since there are two types of cumulative frequency distribution e.g., " less than" and " more than" cumulative frequencies. We can have two types of ogives.

i) 'Less than' Ogive: In 'less than' ogive , the less than cumulative frequencies are plotted against the upper class boundaries of the respective classes. It is an increasing curve having slopes upwards from left to right.

ii) 'More than' Ogive : In more than ogive , the more than cumulative frequencies are plotted against the lower class boundaries of the respective classes. It is decreasing curve and slopes downwards from left to right.

2.3.4 Diagrammatic Presentation of Data

A diagram is a visual form for the presentation of statistical data. They present the data in simple , readily comprehensible form. Diagrammatic presentation is used only for presentation of the data in visual form, whereas graphic presentation of the data can be used for further analysis. There are different forms of diagram e.g., Bar diagram, Sub-divided bar diagram, Multiple bar diagram, Pie diagram and Pictogram.

i) Bar diagram: Bar diagram is most useful for categorical data. A bar is defined as a thick line. Bar diagram is drawn from the frequency distribution table representing the variable on the horizontal axis and the frequency on the vertical

axis. The height of each bar will be corresponding to the frequency or value of the variable.

ii) Sub- divided bar diagram: Study of sub classification of a phenomenon can be done by using sub-divided bar diagram. Corresponding to each sub-category of the data the bar is divided and shaded. There will be as many shades as there will sub portion in a group of data. The portion of the bar occupied by each sub-class reflects its proportion in the total.

iii) Multiple Bar diagram: This diagram is used when comparisons are to be shown between two or more sets of interrelated phenomena or variables. A set of bars for person, place or related phenomena are drawn side by side without any gap. To distinguish between the different bars in a set , different colours , shades are used.

iv) Pie diagram: It is also known as angular diagram. A pie chart or diagram is a circle divided into component sectors corresponding to the frequencies of the variables in the distribution. Each sector will be proportional to the frequency of the variable in the group. A circle represents 360^{O}. So 360^{O} angles is divided in proportion to percentages. The degrees represented by the various component parts of given magnitude can be obtained by using this formula.

After the calculation of the angles for each component, segments are drawn in the circle in succession, corresponding to the angles at the center for each segment. Different segments are shaded with different colours, shades or numbers.

Self Assessment Questions

1) In ‘ less than’ cumulative frequency distribution, which class limit is omitted

   i) upper

   ii) lower

   iii) last

   iv) none of these

2) Differentiate between following components of a statistical table that is “Caption” and “Stub head” “Head note” and “Foot note”.

   .....

   .....

3) Explain the following terms

   i) Histogram,

      .....

   ii) Bar diagram,

      .....

   iii) Frequency polygon, and

      .....

   iv) Pie diagram.

      .....

2.4 SUMMARISATION OF DATA

In the previous section we have discussed about tabulation of the data and its representation in the form of graphical presentation. In research, comparison between two or more series of the same type is needed to find out the trends of variables. For such comparison, tabulation of data is not sufficient and it is further required to investigate the characteristics of data. The frequency distribution of obtained data may differ in two ways, first in measures of central tendency and second, in the extent to which scores are spread over the central value. Both types of differences are the components of summary statistics.

2.4.1 Measures of Central Tendency

It is the middle point of a distribution. Tabulated data provides the data in a systematic order and enhances their understanding. Generally, in any distribution values of the variables tend to cluster around a central value of the distribution. This tendency of the distribution is known as central tendency and measures devised to consider this tendency is know as measures of central tendency. A measure of central tendency is useful if it represents accurately the distribution of scores on which it is based. A good measure of central tendency must possess the following characteristics:

It should be clearly defined- The definition of a measure of central tendency should be clear and unambiguous so that it leads to one and only one information.

It should be readily comprehensible and easy to compute.

It should be based on all observations- A good measure of central tendency should be based on all the values of the distribution of scores.

It should be amenable for further mathematical treatment.

It should be least affected by the fluctuation of sampling.

In Statistics there are three most commonly used measures of central tendency. These are:

1) Arithmetic Mean 2) Median, and 3) Mode

1) Arithmetic Mean: The arithmetic mean is most popular and widely used measure of central tendency. Whenever we refer to the average of data, it means we are talking about its arithmetic mean. This is obtained by dividing the sum of the values of the variable by the number of values. It is also a useful measure for further statistics and comparisons among different data sets. One of the major limitations of arithmetic mean is that it cannot be computed for open-ended class-intervals.

2) Median: Median is the middle most value in a data distribution. It divides the distribution into two equal parts so that exactly one half of the observations is below and one half is above that point. Since median clearly denotes the position of an observation in an array, it is also called a position average. Thus more technically, median of an array of numbers arranged in order of their magnitude is either the middle value or the arithmetic mean of the two middle values. It is not affected by extreme values in the distribution.

3) Mode: Mode is the value in a distribution that corresponds to the maximum concentration of frequencies. It may be regarded as the most typical of a series value. In more simple words, mode is the point in the distribution comprising maximum frequencies therein.

2.4.2 Measures of Dispersion

In the previous section we have discussed about measures of central tendency. By knowing only the mean, median or mode, it is not possible to have a complete picture of a set of data. Average does not tell us about how the score or measurements are arranged in relation to the center. It is possible that two sets of data with equal mean or median may differ in terms of their variability. Therefore, it is essential to know how far these observations are scattered from each other or from the mean. Measures of these variations are known as the 'measures of dispersion'. The most commonly used measures of dispersion are range, average deviation, quartile deviation, variance and standard deviation.

i) Range

Range is one of the simplest measures of dispersion. It is designated by 'R'. The range is defined as the difference between the largest score and the smallest score in the distribution. It gives the two extreme values of the variable. A large value of range indicates greater dispersion while a smaller value indicates lesser dispersion among the scores. Range can be a good measure if the distribution is not much skewed.

ii) Average deviation

Average deviation refers to the arithmetic mean of the differences between each score and the mean. It is always better to find the deviation of the individual observations with reference to a certain value in the series of observation and then take an average of these deviations. This deviation is usually measured from mean or median. Mean, however, is more commonly used for this measurement.

Merits: It is less affected by extreme values as compared to standard deviation. It provides better measure for comparison about the formation of different distributions.

iii) Standard deviation

Standard deviation is the most stable index of variability. In standard deviation, instead of the actual values of the deviations we consider the squares of deviations and the outcome is known as variance. Further, the square root of this variance is known as standard deviation and designated as SD. Thus, standard deviation is the square root of the mean of the squared deviations of the individual observations from the mean. The standard deviation of the sample (ó) and population denoted by (ì) respectively. If all the score have an identical value in a sample, the SD will be 0 (zero).

Merits: It is based on all observations. It is amenable to further mathematical treatments.

Of all measures of dispersion, standard deviation is least affected by fluctuation of sampling.

2.4.3 Skewness and Kurtosis

There are two other important characteristics of frequency distribution that provide useful information about its nature. They are known as skewness and kurtosis.

i) Skewness

Skewness is the degree of asymmetry of the distribution. In some frequency distributions scores are more concentrated at one end of the scale. Such a distribution

is called a skewed distribution. Thus, skewness refers to the extent to which a distribution of data points is concentrated at one end or the other. Skewness and variability are usually related, the more the skewness the greater the variability.

ii) Kurtosis

The term 'kurtosis' refers to the 'peakedness' or flatness of a frequency distribution curve when compared with normal distribution curve. The kurtosis of a distribution is the curvedness or peakedness of the graph.

If a distribution is more peaked than normal it is said to be leptokurtic. This kind of peakedness implies a thin distribution.

On the other hand, if a distribution is more flat than the normal distribution it is known as Platykurtic distribution.

A normal curve is known as mesokurtic.

2.4.4 Advantages and Disadvantages of Descriptive Statistics

The *Advantages* of Descriptive statistics are given below:

- It is essential for arranging and displaying data.
- It forms the basis of rigorous data analysis.
- It is easier to work with, interpret, and discuss than raw data.
- It helps in examining the tendencies, variability, and normality of a data set.
- It can be rendered both graphically and numerically.
- It forms the basis for more advanced statistical methods.

The *disadvantages* of descriptive statistics can be listed as given below:

- It can be misused, misinterpreted, and incomplete.
- It can be of limited use when samples and populations are small.
- It offers little information about causes and effects.
- It can be dangerous if not analysed completely.
- There is a risk of distorting the original data or losing important detail.

Self Assessment Questions

1) Which one of the alternative is appropriate for descriptive statistics?

   i) In a sample of school children, the investigator found an average IQ was 110.

   ii) A class teacher calculates the class average on their final exam. Was 64%.

2) State whether the following statements are *True* (T) or *False* (F).

   i) Mean is affected by extreme values ( )

   ii) Mode is affected by extreme values ( )

iii)	Mode is useful in studying qualitative facts such as intelligence	( )
iv)	Median is not affected by extreme values	( )
v)	Range is most stable measures of variability	( )
vi)	Standard deviation is most suitable measures of dispersion	( )
vii)	Skewness is always positive	( )

2.5 MEANING OF INFERENTIAL STATISTICS

In the previous section we discussed about descriptive statistics, which basically describes some characteristics of data. But the description or definition of the distribution or observations is not the prime objective of any scientific investigation.

Organising and summarising data is only one step in the process of analysing the data. In any scientific investigation either the entire population or a sample is considered for the study.

In most of the scientific investigations a sample, a small portion of the population under investigation, is used for the study. On the basis of the information contained in the sample we try to draw conclusions about the population. This process is known as statistical inference.

Statistical inference is widely applicable in behavioural sciences, especially in psychology. For example, before the Lok Sabha or Vidhan Sabha election process starts or just before the declaration of election results print media and electronic media conduct exit poll to predict the election result. In this process all voters are not included in the survey, only a portion of voters i.e. sample is included to infer about the population. This is called inferential statistics.

Inferential statistics deals with drawing of conclusions about large group of individuals (population) on the basis of observation of a few participants from among them or about the events which are yet to occur on the basis of past events. It provides tools to compute the probabilities of future behaviour of the subjects.

Inferential statistics is the mathematics and logic of how this generalisation from sample to population can be made.

There are two types of inferential procedures: (1) Estimation, (2) Hypothesis testing.

2.5.1 Estimation

In estimation, inference is made about the population characteristics on the basis of what is discovered about the sample. There may be sampling variations because of chance fluctuations, variations in sampling techniques, and other sampling errors. Estimation about population characteristics may be influenced by such factors. Therefore, in estimation the important point is that to what extent our estimate is close to the true value.

Characteristics of Good Estimator: A good statistical estimator should have the following characteristics, (i) Unbiased (ii) Consistent (iii) Accuracy . These are being dealt with in detail below.

i) Unbiased

An unbiased estimator is one in which, if we were to obtain an infinite number of

random samples of a certain size, the mean of the statistic would be equal to the parameter. The sample mean, ($\bar{x}$) is an unbiased estimate of population mean (μ) because if we look at possible random samples of size N from a population, then mean of the sample would be equal to μ.

ii) Consistent

A consistent estimator is one that as the sample size increased, the probability that estimate has a value close to the parameter also increased. Because it is a consistent estimator, a sample mean based on 20 scores has a greater probability of being closer to (μ) than does a sample mean based upon only 5 scores

ii) Accuracy

The sample mean is an unbiased and consistent estimator of population mean (μ). But we should not over look the fact that an estimate is just a rough or approximate calculation. It is unlikely in any estimate that ($\bar{x}$) will be exactly equal to population mean (μ). Whether or not x is a good estimate of (μ) depends upon the representative ness of sample, the sample size, and the variability of scores in the population.

2.5.2 Point Estimation

We have indicated that $\bar{x}$ obtained from a sample is an unbiased and consistent estimator of the population mean (μ). Thus, if an investigator obtains Adjustment Score from 100 students and wanted to estimate the value of (μ) for the population from which these scores were selected, researcher would use the value of $\bar{x}$ as an estimate of population mean (μ). If the obtained value of x were 45.0 then this value would be used as estimate of population mean (μ).

This form of estimate of population parameters from sample statistic is called point estimation. Point estimation is estimating the value of a parameter as a single point, for example, population mean (μ) = 45.0 from the value of the statistic $\bar{x}$ = 45.0

2.5.3 Interval Estimation

A point estimate of the population mean (μ) almost is assured of being in error, the estimate from the sample will not equal to the exact value of the parameter. To gain confidence about the accuracy of this estimate we may also construct an interval of scores that is expected to include the value of the population mean. Such intervals are called confidence interval. A confidence interval is a range of scores that is expected to contain the value of (μ). The lower and upper scores that determine the interval are called confidence limits. A level of confidence can be attached to this estimate so that, the researcher can be 95% or 99% confidence level that encompasses the population mean.

Self Assessment Questions

1) What is statistical inference?

.....

.....

2) Explain with illustrations the concept of

i) Estimation,

.....

ii) Point estimation,

.....

iii) Interval estimation

.....

3) What are the procedures involved in statistical inference?

.....

.....

2.6 HYPOTHESIS TESTING

Inferential statistics is closely tied to the logic of hypothesis testing. In hypothesis testing we have a particular value in mind. We hypothesize that this value characterise the population of observations. The question is whether that hypothesis is reasonable in the light of the evidence from the sample. In estimation no particular population value need be stated. Rather, the question is , what is the population value. For example, Hypothesis testing is one of the important areas of statistical analyses. Sometimes hypothesis testing is referred to as statistical decision-making process. In day-to-day situations we are required to take decisions about the population on the basis of sample information.

2.6.1 Statement of Hypothesis

A statistical hypothesis is defined as a statement, which may or may not be true about the population parameter or about the probability distribution of the parameter that we wish to validate on the basis of sample information.

Most times, experiments are performed with random samples instead of the entire population and inferences drawn from the observed results are then generalised over to the entire population.

But before drawing inferences about the population it should be always kept in mind that the observed results might have come due to chance factor. In order to have an accurate or more precise inference, the chance factor should be ruled out.

The probability of chance occurrence of the observed results is examined by the null hypothesis (H_0). Null hypothesis is a statement of no differences. The other way to state null hypothesis is that the two samples came from the same population. Here, we assume that population is normally distributed and both the groups have equal means and standard deviations.

Since the null hypothesis is a testable proposition, there is counter proposition to it known as alternative hypothesis and denoted by H_1. In contrast to null hypothesis, the alternative hypothesis ($H_{1)}$ proposes that

i) the two samples belong to two different populations,

ii) their means are estimates of two different parametric means of the respective population, and

iii) there is a significant difference between their sample means.

The alternative hypothesis (H_1) is not directly tested statistically; rather its acceptance

or rejection is determined by the rejection or retention of the null hypothesis. The probability 'p' of the null hypothesis being correct is assessed by a statistical test. If probability 'p' is too low, H_0 is rejected and H_1 is accepted.

It is inferred that the observed difference is significant. If probability 'p' is high, H_0 is accepted and it is inferred that the difference is due to the chance factor and not due to the variable factor.

2.6.2 Level of Significance

The level of significance ($p<.05$) is that probability of chance occurrence of observed results up to and below which the probability 'p' of the null hypothesis being correct is considered too low and the results of the experiment are considered significant ($p \leq$).

On the other hand, if p exceeds , the null hypothesis (H_0) cannot be rejected because the probability of it being correct is considered quite high and in such case, observed results are not considered significant ($p >$).

The selection of level of significance depends on the choice of the researcher. Generally level of significance is taken to be 5% or 1%, i.e., = .05 or = .01). If null hypothesis is rejected at .05 level, it means that the results are considered significant so long as the probability 'p' of getting it by mere chance of random sampling works out to be 0.05 or less ($p< .05$). In other words, the results are considered significant if out of 100 such trials only 5 or less number of the times the observed results may arise from the accidental choice in the particular sample by random sampling.

2.6.3 One-tail and Two-tail Test

Depending upon the statement in alternative hypothesis (H_1), either a one-tail or two-tail test is chosen for knowing the statistical significance. A one-tail test is a directional test. It is formulated to find the significance of both the magnitude and the direction (algebraic sign) of the observed difference between two statistics. Thus, in two-tailed tests researcher is interested in testing whether one sample mean is significantly higher (alternatively lower) than the other sample mean.

2.7 ERRORS IN HYPOTHESIS TESTING

In hypothesis testing, there would be no errors in decision making as long as a null hypothesis is rejected when it is false and also a null hypothesis is accepted when it is true. But the decision to accept or reject the null hypothesis is based on sample data. There is no testing procedure that will ensure absolutely correct decision on the basis of sampled data. There are two types of errors regarding decision to accept or to reject a null hypothesis.

2.7.1 Type I Error

When the null hypothesis is true, a decision to reject it is an error and this kind of error is known as type I error in statistics. The probability of making a type I error is denoted as 'á' (read as alpha). The null hypothesis is rejected if the probability 'p' of its being correct does not exceed the p. The higher the chosen level of p for considering the null hypothesis, the greater is the probability of type I error.

2.7.2 Type II Error

When null hypothesis is false, a decision to accept it is known as type II error. The

probability of making a type II error is denoted as 'â' (read as beta). The lower the chosen level of significance p for rejecting the null hypothesis, the higher is the probability of the type II error. With a lowering of p, the rejection region as well as the probability of the type I error declines and the acceptance region (1-p) widens correspondingly.

The goodness of a statistical test is measured by the probability of making a type I or type II error. For a fixed sample size n, á and â are so related that reduction in one causes increase in the other. Therefore, simultaneous reductions in á and â are not possible. If n is increased, it is possible to decrease both á and â.

2.7.3 Power of a Test

The probability of committing type II error is designated by â. Therefore, 1-â is the probability of rejecting null hypothesis when it is false. This probability is known as the power of a statistical test. It measures how well the test is working. The probability of type II error depends upon the true value of the population parameter and sample size n.

Self Assessment Questions

1) Fill in the blanks

   i) Alternative hypothesis is a statement of ..... difference.

   ii) Null hypothesis is denoted by .....

   iii) Alternative hypothesis is ..... directly tested statistically.

   iv) ..... is that probability of chance of occurrence of observed results.

   v) One tail test is a ..... statistical test.

   vi) When the null hypothesis is true, a decision to reject is known as.....

   vii) When a null hypothesis is false, a decision to accept is known as.....

2.8 GENERAL PROCEDURE FOR TESTING A HYPOTHESIS

Step 1. Set up a null hypothesis suitable to the problem.

Step 2. Define the alternative hypothesis.

Step 3. Calculate the suitable test statistics.

Step 4. Define the degrees of freedom for the test situation.

Step 5. Find the probability level 'p' corresponding to the calculated value of the test statistics and its degree of freedom. This can be obtained from the relevant tables.

Step 6. Reject or accept null hypothesis on the basis of tabulated value and calculated value at practical probability level.

There are some situations in which inferential statistics is carried out to test the hypothesis and draw conclusion about the population , for example (i) Test of hypothesis about a population mean (Z test), (ii) Testing hypothesis about a population mean (small sample ' t' test).

2.9 LET US SUM UP

Descriptive statistics are used to describe the basic features of the data in investigation. Such statistics provide summaries about the sample and measures. Data description comprises two operations: organising data and describing data. Organising data includes: classification, tabulation, graphical and diagrammatic presentation of raw scores. Whereas, measures of central tendency and measures of dispersion are used in describing the raw scores.

In the above section, the basic concepts and general procedure involved in inferential statistics are also discussed. Inferential statistics is about inferring or drawing conclusions from the sample to population. This process is known as statistical inference. There are two types of inferential procedures: estimation and hypothesis testing. An estimate of unknown parameter could be either point or interval. Hypothesis is a statement about a parameter. There are two types of hypotheses: null and alternative hypotheses. Important concepts involved in the process of hypothesis testing example, level of significance, one tail test, two tail test, type I error, type II error, power of a test are explained. General procedure for hypothesis testing is also given.

2.10 UNIT END QUESTIONS

1) What is descriptive statistics? Discuss its advantages and disadvantages.

2) What do you mean by organisation of data? State different methods of organising raw data.

3) Define measures of dispersion. Why is it that standard deviation is considered as the best measures of variability?

4) Explain the importance of inferential statistics.

5) Describe the important properties of good estimators.

6) Discuss the different types of hypothesis formulated in hypothesis testing.

7) Discuss the errors involved in hypothesis testing.

8) Explain the various steps involved in hypothesis testing.

2.11 GLOSSARY

Classification : A systematic grouping of data

Cumulative frequency distribution : A classification, which shows the cumulative frequency below, the upper real limit of the corresponding class interval.

Data : Any sort of information that can be analysed.

Discrete data : When data are counted in a classification.

Exclusive classification : The classification system in which the upper limit of the class becomes the lower limit of next class

Frequency distribution : Arrangement of data values according to their magnitude.

Inclusive classification : When the lower limit of a class differs the upper limit of its successive class.

Mean : The ratio between total and numbers of scores.

Median : The mid point of a score distribution.

Mode : The maximum occurring score in a score distribution.

Central Tendency : The tendency of scores to bend towards center of distribution.

Dispersion : The extent to which scores tend to scatter from their mean and from each other.

Standard Deviation : The square root of the sum of squared deviations of scores from their mean.

Skewness : Tendency of scores to polarize on either side of abscissa.

Kurtosis : Curvedness of a frequency distribution graph.

Range : Difference between the two extremes of a score distribution.

Confidence level : It gives the percentage (probability) of samples where the population mean would remain within the confidence interval around the sample mean.

Estimation : It is a method of prediction about parameter value on the basis Statistic.

Hypothesis testing : The statistical procedures for testing hypotheses..

Level of significance : The probability value that forms the boundary between rejecting and not rejecting the null hypothesis.

Null hypothesis : The hypothesis that is tentatively held to be true (symbolised by Ho)

One-tail test : A statistical test in which the alternative hypothesis specifies direction of the departure from what is expected under the null hypothesis.

Parameter : It is a measure of some characteristic of the population.

Population : The entire number of units of research interest.

Power of a test : An index that reflects the probability that a statistical test will correctly reject the null hypothesis relative to the size of the sample involved.

Sample : A sub set of the population under study.

Statistical inference : It is the process of concluding about an unknown population from known sample drawn from it.

Statistical hypothesis : The hypothesis which may or may not be true about the population parameter.

t-test : It is a parametric test for the significance of differences between means.

Type I error : A decision error in which the statistical decision is to reject the null hypothesis when it is actually true.

Type II error : A decision error in which the statistical decision is not to reject the null hypothesis when it is actually false.

Two-tail test : A statistical test in which the alternative hypothesis does not specify the direction of departure from what is expected under the null hypothesis.

2.12 SUGGESTED READINGS

Asthana, H. S. and Bhushan, B. (2007). *Statistics for Social Sciences* (with SPSS Application). Prentice Hall of India, New Delhi.

Garret, H. E. (2005). *Statistics in Psychology and Education*. Jain publishing, India.

Elhance, D. N., and Elhance, V. (1988). *Fundamentals of Statistics*. Kitab Mahal, Allahabad

Nagar, A. L., and Das, R. K. (1983). *Basic Statistics*. Oxford University Press, Delhi.

Sani, F., and Todman, J. (2006). *Experimental Design and Statistics for Psychology.* Blackwell Publishing U.K.

Yale, G. U., and M.G. Kendall (1991). *An Introduction to the Theory of Statistics*. Universal Books, Delhi.

UNIT 3 TYPE I AND TYPE II ERRORS

Structure

3.0 Introduction

3.1 Objectives

3.2 Definition and Concepts

- 3.2.1 Hypothesis Testing
- 3.2.2 The Core Logic of Hypothesis Testing
- 3.2.3 The Hypothesis – Testing Process
- 3.2.4 Implications of Rejecting or Failing to Reject the Null Hypothesis
- 3.2.5 One-Tailed and Two-Tailed Hypothesis Tests
- 3.2.6 Decision Errors

3.3 Type I Error

3.4 Type II Error

3.5 Relationship between Type I and Type II Errors

3.6 Let Us Sum Up

3.7 Unit End Questions

3.8 Glossary

3.9 Suggested Readings

3.0 INTRODUCTION

Each and every discipline needs Statistics and thus is the importance of statistics. One finds that statistics is of great importance to government organisations, non government organisation, experts of all the fields and also to students. Statistics is used for a wide variety of purposes. It is also true that all the time they are not accurate and correct. Sometimes the results are known and sometimes unknown. In the words of Statistics these are known as errors. To achieve accuracy in the concerned field it is important to understand these concepts in detail. It is also important to understand and discuss the related concepts which would be helpful to understand type I and type II errors. In this unit we would be dealing with the definition and concept of errors in statistics and focus on type I and type II errors wich are essential to understand when we deal with statistics and make interpretation of the results using statistics.

3.1 OBJECTIVES

After completing this unit, you will be able to:

- define and differentiate between Type I and Type II errors;
- describe probability concept and the level of significance;
- define and differentiate between one tailed and two tailed tests;
- explain the significance of Normal probability curve;
- define the Cut off sample scores; and
- describe what is z-scores.

3.2 DEFINITION AND CONCEPTS

Before moving onwards we should know the related concepts of Type I and Type II Errors. The concepts that need to be understood include the following:

1) Hypothesis testing

2) The hypothesis – testing process

3) Null Hypothesis

4) Population

5) Sample

6) Rejecting and accepting null hypothesis

7) One-tailed and two-tailed hypothesis

8) Decision errors

3.2.1 Hypothesis Testing

Hypothesis testing has a vital role in psychological measurements. By hypothesis we mean the tentative answer to any questions. Hypothesis testing is a systematic procedure for deciding whether the results of a research study, which examines a sample, support a particular theory or practical innovation, which applies to a population. Hypothesis testing is the central theme in most psychology research.

Hypothesis testing involves grasping ideas that make little sense. Real life psychology research involves samples of many individuals. At the same time there are studies which involve a single individual.

3.2.2 The Core Logic of Hypothesis Testing

There is a standard kind of reasoning researchers use for any hypothesis/testing problem. For this example, it works as follows. Ordinarily, among the population of babies that are not given the specially purified vitamin, the chance of a baby's starting to walk at age 8 months or earlier would be less than 2% Thus, walking at 8 months or earlier is highly unlikely among such babies. But what if the randomly selected sample of one baby in our study does start walking by 8 months? If the specially purified vitamin had no effect on this particular baby's walking age (which means that the baby's walking age should be similar to that of babies that were not given the vitamin), it is highly unlikely (less than a 2% chance) that the particular baby we selected at random would start walking by 8 months. So, if the baby in our study does in fact start walking by 8 months, that allows us to *reject* the idea that the specially purified vitamin has no effect. And if we reject the idea that the specially purified vitamin has no effect, then we must also accept the idea that the specially purified vitamin does have an effect. Using the same reasoning, if the baby starts walking by 8 months, we can reject the idea that this baby comes from a population of babies with a mean walking age of 14 months. We therefore conclude that babies given the specially purified vitamin will start to walk before 14 months. Our explanation for the baby's early walking age in the study is that the specially purified vitamin speeded up the baby's development.

The researchers first spelled out what would have to happen for them to conclude that the special purification procedure makes a differences. Having laid this out in advance the researchers could then go on to carry out their study. In this example,

carrying out the study means giving the specially purified vitamin to a randomly selected baby and watching to see how early that baby walks. Suppose the result of the study is that the baby starts walking before 8 months. The researchers would then conclude that it is unlikely the specially purified vitamin makes no difference and thus also conclude that it does make a difference.

This kind of testing the opposite-of-what-you-predict, roundabout reasoning is at the heart of inferential statistics in psychology. It is something like a double negative. One reason for this approach is that we have the information to figure the probability of getting a particular experimental result if the situation of there being no difference is true. In the purified vitamin example, the researchers know what the probabilities are of babies walking at different ages if the specially purified vitamin does not have any effect. It is the probability of babies walking at various ages that is already known from studies of babies in general – that is, babies who have not received the specially purified vitamin. (Suppose the specially purified vitamin has no effect. In that situation, the age at which babies start walking is the same whether or not they receive the specially purified vitamin.)

Without such a tortuous way of going at the problem, in most cases you could just not do hypothesis testing at all. In almost all psychology research, we base our conclusions on this question: What is the probability of getting our research results. If the opposite of what we are predicting were true? That is, we are usually predicting an effect of some kind. However, we decide on whether there is such an effect by seeing if it is unlikely that there is not such an effect. If it is highly unlikely that we would get our research results if the opposite of what we are predicting were true, that allows us to reject that opposite prediction. If we reject that opposite prediction, we are able to accept our prediction. However, if it is likely that we would get our research results if the opposite of what we are predicting were true, we are not able to reject that opposite prediction. If we are not able to reject that opposite prediction, we are not able to accept our prediction.

3.2.3 The Hypothesis – Testing Process

Let's look at example in this time going over each step in some detail. Along the way, we cover the special terminology of hypothesis-testing. Most important, we introduce five steps of hypothesis testing you use for the rest of the course.

Step 1: Restate the question as a research Hypothesis and Null Hypothesis about the populations

Our researchers are interested in the effects on babies in general (not just this particular baby). That is, the purpose of studying samples is to know about populations thus, it is useful to restate the research question in terms of populations.

In our example, we can think of two populations of babies.

Population 1: Babies who take the specially purified vitamin.

Population 2: Babies who do not take the specially purified vitamin

Population 1 comprises those babies who receive the experimental treatment. In our example, we use a sample of one baby to draw a conclusion about the age that babies in Population 1 start to walk. Population 2 is a kind of comparison baseline of what is already known.

The prediction of our research team is that Population 1 babies (those who take the

specially purified vitamin) will on the average walk earlier than population 2 babies (those who do not take the specially purified vitamin) $\mu_1 < \mu_2$

The opposite of the research hypothesis is that the populations are not different in the way predicted. Under this scenario, population 1 babies (those who take the specially purified vitamin) will on the average not walk earlier than Population 2 babies (those who do not take the specially purified vitamin). That is, this prediction is that there is no difference in when population 1 and Population 2 babies start walking. They start at the same time. A statement like this, about a lack of difference between populations, is the crucial opposite of the research hypothesis. It is called a null hypothesis. It has this name because it states the situation in which there is no difference (the difference is "null") between the between the populations. In symbols, the null hypothesis is $\mu_1 < \mu_2^1$.

The research hypothesis and the null hypothesis are complete opposites: if one is true, the other cannot be. In fact, the research hypothesis is sometimes called the alternative hypothesis – that is, it is the alternative to the null hypothesis. This is a bit ironic. As researchers, we care most about the research hypothesis. But when doing the steps of hypothesis so that we can decide about its alternative (the research hypothesis).

Step 2: Determine the Characteristics of the comparison Distribution

Recall that the overall logic of hypothesis testing involves figuring out the probability of getting a particular result if the null hypothesis is true. Thus, you need to know what the situation would be if the null hypothesis were true. Population 2 we know $\mu = 14, \ \sigma = 3,$ and it is normally distributed. If the null hypothesis is true,

Population 1 and Population 2 are the same – in our example, this would mean Populations 1 and 2 both follow a normal curve, $\mu = 14, \ \sigma = 3$.

In the hypothesis-testing process, you want to find out the probability that you could have gotten a sample score as extreme as what you got (say, a baby walking very early) if your sample were from a population with a distribution of the sort you would have if the null hypothesis were true. Thus, in this book we call this distribution a comparison distribution. (The comparison distribution is sometimes called a statistical model or a sampling distribution – an idea we discuss in Chapter 5.) That is, in the hypothesis-testing process, you compare the actual sample's score to this comparison distribution.

In our vitamin example, the null hypothesis is that there is no difference in walking age between babies that take the specially purified vitamin (Population 1) and babies that do not take the specially purified vitamin (Population 2). The comparison distribution is the distribution for Population 2, since this population represents the walking age of babies if the null hypothesis is true. In later chapters, you will learn about different types of comparison distributions, but the same principle applies in all cases: The comparison distribution is the distribution that represents the population situation if the null hypothesis is true.

Step 3-Determine the Cutoff Sample Score on the comparison Distribution at Which the null hypothesis sould be rejected.

Ideally, before conducting a study, researchers set a target against which they will compare their result – how extreme a sample score they would need to decide against the null hypothesis: that is, how extreme the sample score would have to be

for it to be too unlikely that they could get such an extreme score if the null hypothesis were true. This is called the cutoff sample score. (The cutoff sample score is also known as the critical value.)

Step 4: Determine your sample's Score on the Comparison Distribution

The next step is to carry out the study and get the actual result for your sample. Once you have the results for your sample, you figure the Z score for the sample's raw score based on the population mean and standard deviation of the comparison distribution.

Assume that the researchers did the study and the baby who was given the specially purified vitamin started walking at 6 months. The mean of the comparison distribution to which we are comparing these results is 14 months and the standard deviation is 3 months. That is $\mu = 14,\ \sigma = 3$. Thus, a baby who walks at 6 months is 8 months below the population mean. This puts this baby $2\frac{2}{3}$ standard deviations below the population mean. The Z score for this sample baby on the comparison distribution is thus – 2.67 (Z = [6 - 14]/3=-2.67).

Step 5: Decide Whether to reject the null hypothesis

To decide whether to reject the null hypothesis, you compare your actual sample's Z score (from Step 4) to the cutoff Z score (from Step 3). In our example, the actual result was – 2.67. Let's suppose the researchers had decided in advance that they would reject the null hypothesis if the sample's Z score was below – 2. Since – 2.67 is below – 2, the researchers would reject he null hypothesis.

Or, suppose the researchers had used the more conservative 1% significance level. The needed Z score to reject the null hypothesis would then have been – 2.33 or lower. But, again, the actual Z for the randomly selected baby was – 2.67 (a more extreme score than – 2.33). Thus, even with this more conservative cutoff, they would still reject the null hypothesis.

3.2.4 Implications of Rejecting or Failing to Reject the Null Hypothesis

It is important to emphasise two points about the conclusions you can make from the hypothesis-testing process. First, suppose you reject the null hypothesis. Therefore, your result supports the research hypothesis (as in our example). You would still not say that the results prove the research hypothesis or that the results show that the research hypothesis is true. This would be too strong because the results of research studies are based on probabilities. Specifically, they are based on the probability being low of getting your result if the null hypothesis were true. Proven and true are okay in logic and mathematics, but to use these worlds in conclusions from scientific research is quite unprofessional. (It is okay to use true when speaking hypothetically) – for example, "if this hypothesis were true, then…" – but not when speaking of conclusions about an actual result.) what you do say when you reject the null hypothesis is that the results are statistically significant.

Second, when a result is not extreme enough to reject the null hypothesis, you do not say that the result supports the null hypothesis. You simply say the result is not statistically significant.

A result that is not strong enough to reject the null hypothesis means the study was

inconclusive. The results may not be extreme enough to reject the null hypothesis, but the null hypothesis might still be false (and the research hypothesis true). Suppose in our example that the specially purified vitamin had only a slight but still real effect. In that case, we would not expect to find a baby given the purified vitamin to be walking a lot earlier than babies in general. Thus, we would not be able to reject the null hypothesis, even though it is false. (You will learn more about such situations in the Decision Errors section later in this chapter).

Showing the null hypothesis to be true would mean showing that there is absolutely no difference between the populations it is always possible that there is a difference between the populations, but that the difference is much smaller than what the particular study was able to detect. Therefore, when a result is not extreme enough to reject the null hypothesis, the results are inconclusive. Sometimes, however, if studies have been done using large samples and accurate measuring procedures, evidence may build up in support of something close to the null hypothesis – that there is at most very little difference between the populations.

3.2.5 One-Tailed and Two-Tailed Hypothesis Tests

In our examples so far, the researchers were interested in only one direction of result. In our first example, researchers tested whether babies given the specially purified vitamin would walk earlier than babies in general. In the happiness example, the personality psychologists predicted the person who received $10 million would be happier than other people. The researchers in these studies were not interested in the possibility that giving the specially purified vitamin would cause babies to start walking later or that people getting $10 million might become less happy.

Directional hypotheses and One-Tailed tests

The purified vitamin and happiness studies are examples of testing directional hypotheses. Both studies focused on a specific direction of effect. When a researcher makes a directional hypothesis, the null hypothesis is also, in a sense, directional. Suppose the research hypothesis is that getting $10 million will make a person happier. The null hypothesis, then, is that the money will either have no effect or make the person less happy (in symbols, if the research hypothesis is $\mu > \mu_2$, then the null hypothesis is $\mu_1 \leq \mu_2 \leq$ is the symbol for less than or equal to.) thus to reject the null hypothesis, the sample had to have a score in one particular tail of the comparison distribution – the upper extreme or tail (in this example, the top 5%) of the comparison distribution. (When it comes to rejecting the null hypothesis with a directional hypothesis, a score at the other tail would be the same as a score in the middle – that is, it would not allow you to reject the null hypothesis). For this reason, the test of a directional hypothesis is called a one-tailed test. A one-tailed test can be one-tailed in either direction. In the happiness study example, the tail for the predicted effect was at the high end. In the baby study example, the tail for the predicted effect was at the low end (that is, the prediction tested was that babies given the specially purified vitamin would start walking unusually early).

Non-directional hypotheses and two-tailed tests

Sometimes, a research hypothesis states that an experimental procedure will have an effect, without saying whether it will produce a very high score or a very low score. Suppose an organisational psychologist is interested in how a new social skills program will affect productivity. The program could improve productivity by making the working environment more pleasant. Or, the program could hurt productivity by encouraging

people to socialise instead of work. The research hypothesis is that the social skills program changes the level of productivity; the null hypothesis is that the program does not change productivity one way or the other. In symbols, the research hypothesis is $ì_1 \neq ì_2$ the null hypothesis is $ì_1 = ì_2$

When a research hypothesis predicts an effect but does not predict a particular direction for the effect, it is called a non-directional hypothesis. To test the significance of a non-directional hypothesis, you have to take into account the possibility chance of non-directional hypothesis, you have to take into account the possibility that the sample could be extreme at either tail of the comparison distribution. Thus this is called a two-tailed test.

3.2.6 Decision Errors

Another crucial topic for making sense of statistical significance is the kind of errors that are possible in the hypothesis-testing process. The kind of errors we consider here are about how, in spite of doing all your figuring correctly, your consider here are about how, in spite of doing all your figuring correctly, your conclusions from hypothesis-testing can still be incorrect. It is not about making mistakes in calculations or even about using the wrong procedures. That is, mistakes in calculations or even about using the wrong procedures. That is, decision errors are situations in which the right procedures lead to the wrong decisions.

Decision errors are possible in hypothesis testing because you are making decisions about populations based on information in samples. The whole hypothesis testing process is based on probabilities. The hypothesis-testing process is set up to make the probability of decision errors as small as possible. For example, we only decide to reject the null hypothesis if a sample's mean is so extreme that there is a very small probability (say, less than 5%) that we could have gotten such an extreme sample if the null hypothesis is true. But a very small probability is not the same as a zero probability! Thus, in spite of your best intentions, decision errors are always II errors.

3.3 TYPE I ERROR

You make a Type I error if you reject the null hypothesis when in fact the null hypothesis is true. Or, to put it in terms of the research hypothesis, you make a Type I error when you conclude that the study supports the research hypothesis when in reality the research hypothesis is false.

Suppose you carried out a study in which you had set the significance level cut off at a very lenient probability level, such as 20%. This would mean that it would not take a very extreme result to reject the null hypothesis. If you did many studies like this, you would often (about 20% of the time) be deciding to consider the research hypothesis supported when you should not. That is, you would have a 20% chance of making a Type I error.

Even when you set the probability at the conventional .05 or .01 levels, you will still make a Type I error sometimes (5% or 1% of the time). Consider again the example of giving the new therapy to a depressed patient. Suppose the new therapy is not more effective than the usual therapy. However, in randomly picking a sample of one depressed patient to study, the clinical psychologists might just happen to pick a patient whose depression would respond equally well to the new therapy and the usual therapy. Randomly selecting a sample patient like this is unlikely, but such extreme samples are possible, and should this happen, the clinical psychologists

would reject the null hypothesis and conclude that the new therapy is different than the usual therapy. Their decision to reject the null hypothesis would be wrong – a Type I error. Of course, the researchers could not know they had made a decision error of this kind. What reassures researchers is that they know from the logic of hypothesis testing that the probability of making such a decision error is kept low (less than 5% if you use the .05 significance level).

Still, the fact that Type I errors can happen at all is of serious concern to psychologists, who might construct entire theories and research programs, not to mention practical applications, based on a conclusion from hypothesis testing that is in fact mistaken. It is because these errors are of such serious concern that they are called Type I.

As we have noted, researchers cannot tell when they have made a Type I error. However, they can try to carry out studies so that the chance of making a Type I error is as small as possible.

What is the chance of making a Type I error? It is the same as the significance level you set. If you set the significance level at $p < .05$, you are saying you will reject the null hypothesis if there is less than a 5% (.05) chance that you could have gotten your result if the null hypothesis were true. When rejecting the null hypothesis in this way, you are allowing up to a 5% chance that you got your results even though the null hypothesis was actually true. That is, you are allowing a 5% chance of a Type I error.

The significance level, which is the chance of making a Type I error, is called alpha (the Greek letter Ü). The lower the alpha, the smaller the chance of a Type I error. Researchers who do not want to take a lot of risk set alpha lower than .05 such a $p<.001$ in this way the result of a study has to be very extreme in order for the hypothesis testing process to reject the null hypothesis.

Using a .001 significance level is like buying insurance against making a Type I error. However, when buying insurance, the better the protection, the higher the cost. There is a cost in setting the significance level at too extreme a level. We turn to that cost next.

3.4 TYPE II ERROR

If you set a very stringent significance level, such as .001, you run a different kind of risk. With a very stringent significance level, you may carry out a study in which in reality the research hypothesis is true, but the result does not come out extreme enough to reject the null hypothesis. Thus, the decision error you would make is in not rejecting the null hypothesis when in reality the null hypothesis is false to put this in terms of the research hypothesis, you make this kind of decision error when the hypothesis-testing procedure leads you to decide that the results of the study are inconclusive when in reality the research hypothesis is true. This is called a Type II error. The probability of making a Type II error is called beta (the Greek letter β).

3.5 RELATIONSHIP BETWEEN TYPE I AND TYPE II ERRORS

When it comes to setting significance levels, protecting against one kind of decision error increases the chance of making the other. The insurance policy against Type I error (setting a significance level of, say, .001) has the cost of increasing the chance of making a Type II error. (This is because with a stringent significance level like

.001, even if the research hypothesis is true, the results have to be quite strong to be extreme enough to reject the null hypothesis.) The insurance policy against Type II error (setting a significance level of say .20) has the cost of increasing the chance of making a Type I error. (This is because with a level of significance like .20, even if the null hypothesis is true, it is fairly easy to get a significant result just by accidentally getting a sample that is higher or lower than the general population before doing the study.)

3.6 LET US SUM UP

Hypothesis testing considers the probability that the result of a study could have come about even if the experimental procedure had no effect. If this probability is low, the scenario of no effect is rejected and the theory behind the experimental procedure is supported.

The expectation of an effect is the research hypothesis, and the hypothetical situation of no effect is the null hypothesis.

When a result (that is, a sample score) is so extreme that the result would be very unlikely if the null hypothesis were true, the null hypothesis is reject and the research hypothesis supported. If the result is not that extreme, the null hypothesis is not rejected and the study is inconclusive.

Psychologists usually consider a result too extreme if it is less likely than 5% (that is, a significance level of .05) to have come about, if the null hypothesis were true. Psychologists sometimes use a more stringent 1% (.01 significance level), or even .01% (.001) significance level), cutoff.

The cutoff percentage is the probability of the result being extreme in a predicted direction in a directional or one-tailed test. The cutoff percentages are the probability of the result being extreme in either direction in a non-directional or two-tailed test.

There are two kinds of decision errors one can make in hypothesis testing. A Type I error is when a researcher rejects the null hypothesis, but the null hypothesis is actually true. A Type II error is when a researcher does not reject the null hypothesis, but the null hypothesis is actually false.

There has been much controversy about significance tests, including critiques of the basic logic and, especially, that they are often misused. One major way significance tests are misused is when researchers interpret not rejecting the null hypothesis as demonstrating that the null hypothesis is true.

Research articles typically report the results of hypothesis testing by saying a result was or was not significant and giving the probability level cutoff (usually 5% or 1%) the decision was based on. Research articles rarely mention decision errors.

3.7 UNIT END QUESTIONS

1) Fill in the blanks with appropriate terms:

   i) The research hypothesis and_____are completely opposite.

   ii) Cutoff sample score are also known as_____

   iii) In Type I Error we _____ hypothesis when it is true.

   iv) The hypothesis in which we Accept Null hypothesis is called_____

2) Mark (T) for True statement and (F) for False Statement:

   i) The probability of making Type II error is called. ( )

   ii) The significance level, which is chance of making Type II Error, is called. ( )

   iii) The significance level, which is chance of making Type I error is called. ( )

   iv) One directional hypothesis is termed as two tailed test. ( )

   v) A hypothesis which predicts an effect but does not predict a particular direction for the effect is called non-directional hypothesis. ( )

   vi) One tailed test has either direction only. ( )

3) Give brief answers of the following questions:

   i) What is Comparison distribution?

   ii) What is research hypothesis?

   iii) What is directional hypothesis?

   iv) What is Type I error?

   v) What do you mean by two tailed test?

   vi) Describe briefly the relationship between Type I & Type II errors.

3.8 GLOSSARY

Hypothesis : Tentative statement which can be tested.

Research hypothesis : Statement about the predicted relation between populations.

Null hypothesis : A Statement opposite to the research hypothesis.

Alternate hypothesis : A statement which is opposite to the null hypothesis

Level of Significance : Probability of getting statistical significance of null hypothesis is accurately true.

Comparison distributions : Distribution used in hypothesis testing.

One tailed test : Hypothesis testing procedure for a directional hypothesis

Two tailed test : Hypothesis testing procedure for a non-directional hypothesis

Sample : Scores of particular group of people studied.

Type I Error : When we reject a null hypothesis when it is true

Type II	:	When we accept a null hypothesis when it is false
α(alpha)	:	Probability of making type – I Error
β(Beta)	:	Sampling distribution Probability of making type I Error
Normal curve	:	Bell shaped frequency distribution that is Symmetrical and unimodel.

3.9 SUGGESTED READINGS

Asthana H.S, and Bhushan. B. (2007) *Statistics for Social Sciences* (with SPSS Applications).

B.L. Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Gupta, S.C. (1990) *Fundamentals of Statistics*

Sidney Siegel, & N. John Castetellan, Jr. *Non-parametric Statistics for the Behavioural Science*. McGraw Hill Books company

Y.P. Aggarwal. *Statistical Methods Concepts*, Application & Computation Sterling homosedasity publishers Pvt. Ltd.

UNIT 4 SETTING UP THE LEVELS OF SIGNIFICANCE

Structure

4.0 Introduction

4.1 Objectives

4.2 Hypothesis Testing

4.3 Null Hypothesis

4.4 Errors in Hypothesis Testing

- 4.4.1 Basic Experimental Situations in Hypothesis Testing

4.5 Confidence Limits

- 4.5.1 Meaning and Concept of Level of Significance
- 4.5.2 Application and Interpretation of Standard Error of the Mean in Small Samples
- 4.5.3 The Standard Error of a Median, σ_{Mdn}

4.6 Setting up Level of Confidence or Significance

- 4.6.1 Size of the Sample
- 4.6.2 Two-tailed and One-tailed Tests of Significance
- 4.6.3 One Tailed Test

4.7 Steps in Setting up the Level of Significance

- 4.7.1 Formulating Hypothesis and Stating Conclusions
- 4.7.2 Types of Errors for a Hypothesis Testing

4.8 Let Us Sum Up

4.9 Unit End Questions

4.10 Glossary

4.11 Suggested Readings

4.0 INTRODUCTION

In behavioural sciences nothing is absolute. Therefore, while obtaining the findings through statistical analyses, behavioural scientists usually ignore the error to the maximum of 5%. In statistics, a result is called statistically significant if it is unlikely to have occurred by chance. The phrase, "test of significance" was coined by Ronald Fisher. As used in statistics, significance does not mean importance or meaning fitness as it does in everyday speech. In this unit we will be dealing with the definition and concept of level of significance and how the level of significance is decided. Since level of significance is related to hypothesis testing we will be dealing with null hypothesis and alternative hypothesis and how these are to be tested in different types of experiments. While dealing with hypothesis testing we will also be dealing with experimental designs , errors in hypothesis testing etc. We will also learn what is meant by confidence limits and how these are established.

4.1 OBJECTIVES

After completing this unit, you will be able to:

- define and put forward the concept of null hypothesis;

- describe the process of hypothesis testing;
- explain the confidence limits;
- elucidate the errors in hypothesis testing and its relationship to levels of significance;
- explain level of significance;
- describe the setting up of level of significance; and
- analyse the experimental designs in relation to levels of significance.

4.2 HYPOTHESIS TESTING

Many a time, we strongly believe some results to be true. But after taking a sample, we notice that one sample data does not wholly support the result. The difference is due to

i) the original belief being wrong, or

ii) the sample being slightly one sided.

Tests are, therefore, needed to distinguish between two possibilities. These tests tell about the likely possibilities and reveal whether or not the difference can be due to only chance elements. If the difference is not due to chance elements, it is significant and, therefore, these tests are called tests of significance. The whole procedure is known as *Testing of Hypothesis*.

Setting up and testing hypotheses are essential part of statistical inference. In order to formulate such a test, usually some theory is put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. For example, the hypothesis may be the claim that a new drug is better than the current drug for treatment of a disease, diagnosed through a set of symptoms.

In each problem considered, the question of interest is simplified into two competing claims that is the hypotheses between which we have a choice; the null hypothesis, denoted by Ho, against the alternative hypothesis, denoted by H_1. These two competing claims or hypotheses are not however treated on an equal basis. Special consideration is given to the null hypothesis which states that there is no difference between the two drugs. Null hypothesis is also called as "No Difference" hypothesis. We have two common situations: Let us say we have formulated the null hypothesis stating that "There will be no difference between the two drugs in regard to treating a disorder" We carry out an experiment to prove that the null hypothesis is true or reject that null hypothesis as the experimental results show that there is a difference in the treatment by the two drugs. Thus we have two situations as given below:

i) The experiment has been carried out in an attempt to prove or reject a particular hypothesis, the null hypothesis. We give priority to the null hypothesis and say that we will reject it only if the evidence against it is sufficiently strong.

ii) If one of the two hypotheses is 'simpler', we give it priority so that a more 'complicated' theory is not adopted unless there is sufficient evidence against the simpler one. For example, it is 'simpler' to claim that there is no difference in the treatment between the two drugs than to state the alternate hypothesis that drug A will be better than dug B, that is stating that there is a difference.

The hypotheses are often statements about population parameters like expected value and variance. Take for example another null hypothesis, H_0, the statement that

the expected value of the height of ten year old boys in the Indian population is not different from that of ten year old girls.

A hypothesis can also be a statement about the distributional form of a characteristic of interest. For instance, the statement that the height of ten year old boys is normally distributed within the Indian population. This is a hypothesis in statement form regarding the distribution of 10 year old boys height in the population.

4.3 NULL HYPOTHESIS

The null hypothesis, H_0, represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. For example, in respect of a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug. We would write H_0: there is no difference between the two drugs on average.

We give special consideration to the null hypothesis. This is due to the fact that the null hypothesis relates to the statement being tested, whereas the alternative hypothesis relates to the statement to be accepted if when the null hypothesis is rejected.

The alternative hypothesis, H_1, is a statement of what a statistical hypothesis test is set up to establish. For example, in a clinical trial of a new drug, the alternative hypothesis might be that the new drug has a different effect, on average, compared to that of the current drug.

We would write H_1: the two drugs have different effects, on average.

The alternative hypothesis might also be that the new drug is better, on average, than the current drug. In this case, we would write H_1: the new drug is better than the current drug, on average.

The final conclusion once the test has been carried out is always given in terms of the null hypothesis. We either 'reject H_0 in favour of H_1' or 'do not reject H_o'; we never conclude 'reject H_1', or even 'accept H_1'.

If we conclude 'do not reject H_0', this does not necessarily mean that the null hypothesis is true, It only suggests that there is not sufficient evidence against H0 in favour of H_1 that we reject the null hypothesis. It only suggests that the alternative hypothesis may be true.

Thus one may state that Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution.

First, a tentative assumption is made about the parameter or distribution. This assumption is called the null hypothesis and is denoted by H_0. An alternative hypothesis (denoted by H_1), which is the opposite of what is stated in the null hypothesis, is then defined. The hypothesis-testing procedure involves using sample data to determine whether or not H_0 can be rejected. If H_0 is rejected, the statistical conclusion is that the alternative hypothesis H_1 is true.

A hypothesis is a statement supposed to be true till it is proved false. It may be based on previous experience or may be derived theoretically. First a statistician or the investigator forms a research hypothesis that an exception is to be tested. Then she/he derives a statement which is opposite to the research hypothesis (noting as H_0). The approach here is to set up an assumption that there is no contradiction between

the believed result and the sample result and that the difference, therefore, can be ascribed solely to chance. Such a hypothesis is called a null hypothesis (H_0). It is the null hypothesis that is actually tested, not the research hypothesis. The object of the test is to see whether the null hypothesis should be rejected or accepted.

If the null hypothesis is rejected, that is taken as evidence in favour of the research hypothesis which is called as the alternative hypothesis (denoted by H_1). In usual practice we do not say that the research hypothesis has been "proved" only that it has been supported.

For example, assume that a radio station selects the music it plays based on the assumption that the average age of its listening audience is 30 years. To determine whether this assumption is valid, a hypothesis test could be conducted with the null hypothesis as H_0: = 30 and the alternative hypothesis as H_1: 30. Based on a sample of individuals from the listening audience, the sample mean age, can be computed and used to determine whether there is sufficient statistical evidence to reject H_0. Conceptually, a value of the sample mean that is "close" to 30 is consistent with the null hypothesis, while a value of the sample mean that is "not close" to 30 provides support for the alternative hypothesis.

Self Assessment Questions

Fill in blanks with appropriate terms.

1) Generally the .05 and the _____ levels of significance are mostly used.

2) Standard Error of mean is calculated by the formula _____.

3) In case of two-tailed test (+1.96) with fall on _____ of the normal curve.

4) In case of .05 level of significance amount of confidence will be _____.

4.4 ERRORS IN HYPOTHESIS TESTING

Ideally, the hypothesis-testing procedure leads to the acceptance of H_0 when H_0 is true and the rejection of H_0 when H_0 is false. Unfortunately, since hypothesis tests are based on sample information, the possibility of errors must be considered. A Type-I error corresponds to rejecting H_0 when H_0 is actually true, and a Type-II error corresponds to accepting H_0 when H_0 is false.

In testing any hypothesis, we get only two results: either we accept or we reject it. We do not know whether it is true or false. Hence four possibilities may arise.

i) The hypothesis is true but test rejects it (Type-I error).

ii) The hypothesis is false but test accepts it (Type-II error).

iii) The hypothesis is true and test accepts it (correct decision).

iv) The hypothesis is false and test rejects it (correct decision)

Type-I Error

In a hypothesis test, a Type-I error occurs when the null hypothesis is rejected when it is in fact true. That is, H0 is wrongly rejected. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average,

than the current drug. That is, there is no difference between the two drugs on average. A Type-I error would occur if we concluded that the two drugs produced different effects when in fact there was no difference between them.

A Type-I error is often considered to be more serious, and therefore more important to avoid, than a Type-TI error.

The hypothesis test procedure is therefore adjusted so that there is a guaranteed 'low' probability of rejecting the null hypothesis wrongly;

This probability is never 0. This probability of a Type-I error can be precisely computed as, P (Type-I error) significance level =

The exact probability of a Type-I error is generally unknown.

If we do not reject the null hypothesis, it may still be false (a Type-I error) as the sample may not be big enough to identify the falseness of the null hypothesis (especially if the truth is very close to hypothesis).

For any given set of data, Type-I and Type-Il errors are inversely related; the smaller the risk of one, the higher the risk of the other.

A Type-I error can also be referred to as an error of the first kind.

Type-II Error

In a hypothesis test, a Type-II error occurs when the null hypothesis, Ho, is not rejected when it is in fact false. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is Ho: there is no difference between the two drugs on average.

A Type-II error would occur if it was concluded that the two drugs produced the same effect, that is, there is no difference between the two drugs on average, when in fact they produced different effects.

A Type-II error is frequently due to sample sizes being too small.

The probability of a Type-II error is symbolised by â and written:

P (Type-II error) = â (but is generally unknown).

A Type-II error can also be referred to as an error of the second kind.

Hypothesis testing refers to the process of using statistical analysis to determine if the observed differences between two or more samples are due to random chance factor (as stated in the null hypothesis) or is it due to true differences in the samples (as stated in the alternate hypothesis).

A null hypothesis (H_0) is a stated assumption that there is no difference in parameters (mean, variance) for two or more populations. The alternate hypothesis (H_1) is a statement that the observed difference or relationship between two populations is real and not the result of chance or an error in sampling.

Hypothesis testing is the process of using a variety of statistical tools to analyse data and, ultimately, to fail to reject or reject the null hypothesis. From a practical point of view, finding statistical evidence that the null hypothesis is false allows you to reject the null hypothesis and accept the alternate hypothesis.

Because of the difficulty involved in observing every individual in a population for

research purposes, researchers normally collect data from a sample and then use the sample data to help answer questions about the population.

A hypothesis test is a statistical method that uses sample data to evaluate a hypothesis about a population parameter.

The hypothesis testing is standard and it follows a specific order as given below.

i) first state a hypothesis about a population (a population parameter, e.g. mean

ii) obtain a random sample from the population and also find its mean , and

iii) compare the sample data with the hypothesis on the scale (standard z or normal distribution).

Self Assessment Questions

1) What is Type I error? Give suitable examples.

.....

.....

2) What is Type II error? Give example.

.....

.....

3) What is hypothesis testing? What are the steps for the same?

.....

.....

4) What is Null hypothesis and alternate hypothesis?

.....

.....

4.4.1 Basic Experimental Situations for Hypothesis Testing

i) It is assumed that the mean, t, is known before treatment. The purpose of the experiment is to determine whether or not the treatment has an effect on the population mean, e.g. a researcher will like to find out whether increased stimulation of infants has an effect on their weight. It is known from national statistics that the mean weight, t, of 2-year old children is 13 kg. The distribution is normal with the standard deviation, á =2 kg.

ii) To test the truth of the claim a researcher may take 16 new born infants and give their parents detailed instructions for giving these infants increased handling and stimulations. At age 2, each of the 16 children will be weighed and the mean weight for the sample will be computed.

iii) The researcher may conclude that the increased handling and stimulation had an effect on the weight of the children if there is a substantial difference in the weights from the population mean.

A hypothesis test is typically used in the context of a research study, i.e. a researcher completes one round of a field investigation and then uses a hypothesis test to

evaluate the results. Depending on the type of research and the type of data, the details will differ from one research situation to another.

The probability of making a Type-I error is denoted by á, and the probability of making a Type-II error is denoted by â.

In using the hypothesis-testing procedure to determine if the null hypothesis should be rejected, the person conducting the hypothesis test specifies the maximum allowable probability of making a Type-I error, called the level of significance for the test.

Common choices for the level of significance are á = 0.05 and á = 0.01. Although most applications of hypothesis testing control the probability of making a Type I error, they do not always control the probability of making a Type-Il error.

A concept known as the p-value provides a convenient basis for drawing conclusions in hypothesis-testing applications. The p-value is a measure of how likely the sample results are, assuming that the null hypothesis is true. The smaller the p-value, the lesser likely are the sample results reliable. If the p-value is less than á, the null hypothesis can be rejected, otherwise, the null hypothesis cannot be rejected. The p-value is often called the observed level of significance for the test.

A hypothesis test can be performed on parameters of one or more populations as well as in a variety of other situations. In each instance, the process begins with the formulation of null and alternative hypotheses about the population. In addition to the population mean, hypothesis-testing procedures are available for population parameters such as proportions, variances, standard deviations, and medians.

Hypothesis tests are also conducted in regression and correlation analysis to determine if the regression relationship and the correlation coefficient are statistically significant.

A goodness-of-fit test refers to a hypothesis test in which the null hypothesis is that the population has a specific probability distribution, such as a normal probability distribution. Non-parametric statistical methods also involve a variety of hypothesis testing procedures.

For example, if it is assumed that the mean of the weights of the population of a college is 55 kg, then the null hypothesis will be: the mean of the population is 55 kg, i.e. Ho: = 55 kg (Null hypothesis). In terms of alternative hypothesis (i) H_1: > 55 kg, (ii) H_1: <55 kg.

Now fixing the limits totally depends upon the accuracy desired. Generally the limits are fixed such that the probability that the difference will exceed the limits is 0.05 or 0.01. These levels are known as the 'levels of significance' and are expressed as 5% or 1% levels of significance.

What does this actually mean? When we say the limits not to exceed .05 level, that means what ever result we get we can say with 95% confidence that these are genuine results and not because of any chance factor. If we say .01 level, then it means that we can say with 99% confidence that the obtained results are genuine and not due to any chance factor.

Rejection of null hypothesis does not mean that the hypothesis is disproved. It simply means that the sample value does not support the hypothesis. Also, acceptance does not mean that the hypothesis is proved. It means simply it is being supported.

Self Assessment Questions

1) What is a goodness of fit test?

.....

.....

2) How do we fix the limits for significance in hyopothesis testing?

.....

.....

3) What are the basis experimental situations for hypothesis testing?

.....

.....

4.5 CONFIDENCE LIMITS

The limits (or range) within which the hypothesis should lie with specified probabilities are called the confidence limits or fiduciary limits. It is customary to take these limits as 5% or 1% levels of significance. If sample value lies between the confidence limits, the hypothesis is accepted; if it does not, the hypothesis is rejected at the specified level of significance.

4.5.1 Meaning and Concept of Levels of Significance

Experimenters and researchers have selected some arbitrary standards—called levels of significance to serve as the cut-off points or critical points along the probability scale, so as to separate the significant difference from the non significant difference between the two statistics, like means or SD's.

Generally, the .05 and the .01 levels of significance are the most popular in social sciences research. The confidence with which an experimenter rejects—or retains—a null hypothesis depends upon the level of significance adopted. These may, hence, sometime be termed as levels of confidence. Their meanings may be clear from the following:

Meaning of Levels of Confidence

Level	Amount of confidence	Interpretation
.05	95%	If the experiment is repeated a 100 times, only on five occasions the obtained mean will fall outside the limited $\mu \pm 1.96$ SE
.01	99%	If the experiment is repeated a 100 times, only on one occasions the obtained mean will fall outside the limited $\mu \pm 2.58$ SE

The values 1.96 and 2.58 have been taken from the t tables keeping large samples in view. The .01 level is more rigorous and higher a standard as compared to the .05 level and would require a larger value of the critical ratio for the rejection of the

Ho. Hence if an obtained value of t is significant at 01 level, it is automatically significant at .05 level but the reverse is not always true.

4.5.2 Application and Interpretation of Standard Error of the Mean (SEM) in Small Samples

The procedure of calculation and interpretation of Standard Error of Mean in small samples differs from that for large samples, in two respects.

1) The denominator N-1 instead of N is used in the formula for calculation of the SD of the sample.

2) The appropriate distribution to be used for small samples is t distribution instead of normal distribution.

The rest of the line of reasoning used in determining and interpreting SE in small samples is similar to that for the large samples.

4.5.3 The Standard Error of a Median, σ_{Mdn}

It has been established that the variability of the sample medians is about 25 per cent greater than the variability of means in a normally distributed population. Hence the standard error of a median can be estimated by using the formulas:

$$SE_{Mdn,}\ ó_{Mdn} = \frac{1.253\sigma}{\sqrt{N}}$$

$$ó_{Mdn} = \frac{1.253\sigma}{\sqrt{N}}$$

(Standard Error of the Median in terms of ó and Q)

Self Assessment Questions

1) Describe confidence limits.

.....

.....

2) Elucidate the concept of significance level.

.....

.....

3) What is standard error of the mean? How is it useful in hypothesis testing?

.....

.....

4) What is standard error of median? How is it calculated ? What is its significance**?**

.....

.....

4.6 SETTING UP THE LEVEL OF CONFIDENCE OR SIGNIFICANCE

The experimenter has to take a decision about the level of confidence or significance at which the hypothesis is going to be tested. At times the researcher may decide to use 0:05 or 5% level of significance for rejecting a null hypothesis (when a hypothesis is rejected at the 5% level it is said that the chances are 95 out of 100, that the hypothesis is not true and only 5 chances out of 100 that it is true). At other times, the researcher may prefer to make it more rigid and therefore, use the 0.01 or 1% level of significance. If a hypothesis is rejected at this level, the chances are 99 out of 100, that the hypothesis is not true and that only 1 chance out of 100 is true. This level on which we reject the null hypothesis, is established before doing the actual experiment (before collecting data). Later we have to adhere to it.

4.6.1 Size of the Sample

The sampling distribution of the differences between means may look like a normal curve or *t* distribution curve depending upon the size of the samples drawn from the population. The t distribution is a theoretical probability distribution. It is symmetrical, bell-shaped, and similar to the standard normal curve. It differs from the standard normal curve, however, in that it has an additional parameter, called degrees of freedom, which changes its shape.

If the samples are large (N = 30 or greater than 30), then the distribution of differences between means will be a normal one. If it is small (N is less than 30), then the distribution will take the form of a I distribution and the shape of the t-curve will vary with the number of degrees of freedom.

In this way, for large samples, statistics advocating normal distribution of the characteristics in the given population will be employed, while for small samples, the small sample statistics will be used.

Hence in the case of large samples possessing a normal distribution of the differences of means, the value of standard error used to determine the significance of the difference between means will be in terms of standard sigma (z) scores. On the other hand, in the case of small samples possessing a t- distribution of differences between means, we will make use of t values rather than z scores of the normal curve. From the normal curve table we see that 95% and 99% cases lie at the distance of l.96 and 2.58 . Therefore, the sigma or z scores of 1.96 and 2.58 are taken as critical values for rejecting a null hypothesis.

If a computed z value of the standard error of the differences between means approaches or exceeds the values 1.96 and 2.58, then we may safely reject a null hypothesis at the 0.05 and 0.01 levels.

To test the null hypothesis in the case of small sample means, we first compute the t ratio in the same manner as z scores in case of large samples. Then we enter the table of t distribution (Table C in the Appendix) with $N_1 + N_2 — 2$ degrees of freedom and read the values of t given against the row of $N_1 + N_2 — 2$ degrees of freedom and columns headed by 0.05 and 0.01 levels of significance. If our computed t ratio approaches or exceeds the values of t read from the table, we will reject the established null hypothesis at the 0.05 and 0.01 levels of significance, respectively.

4.6.2 Two-Tailed and One-Tailed Tests of Significance

Two-tailed test. In making use of the two-tailed test for determining the significance of the difference between two means, we should know whether or not such a difference between two means really exists and how trustworthy and dependable this difference is.

In all such cases, we merely try to find out if there is a significant difference between two sample means; whether the first mean is larger or smaller than the second, is of no concern. We do not care for the direction of such a difference, whether positive or negative. All that we are interested in, is a difference.

Consequently, when an experimenter wishes to test the null hypothesis, Ho: $M_1 = 0$, against its possible rejection and finds that it is rejected, then the researcher may conclude that a difference really exists between the two means, however no assertion is made about the direction of the difference. Such a test is a non-directional test.

It is also named as *two-tailed test*, because it employs both sides, positive and negative, of the distribution (normal or t distribution) in the estimation of probabilities. Let us consider the probability at 5% significance level in a two-tailed test with the help of figure given below:

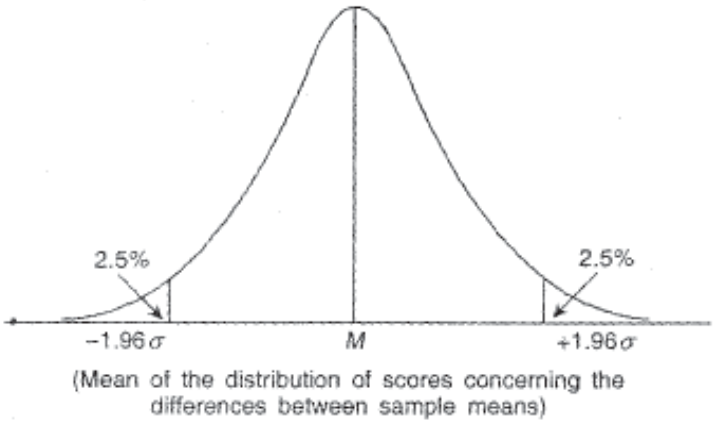


Fig.: Two-tailed test at the 5% level

(Mean of the distribution of scores concerning the differences between sample means)

Two-tailed test at the 5% level.

Therefore, while using both the tails of the distribution we can say that area of the normal curve falls to the right of 1.96 standard deviation units above the mean and 2.5% falls to the left of 1.96 standard deviation units below the mean.

The area outside these limits is 5% of the total area under the curve. In this way, for testing the significance at the 5% level, we may reject a null hypothesis if the computed error of the difference between means reaches or exceeds the yardstick 1.96.

Similarly, we may find that a value of 2.58 is required to test the significance at the 1% level in the case of a two-tailed test.

4.6.3 One-tailed Test

As we have seen, a two-tailed or a non-directional test is appropriate, if we are only concerned with the absolute magnitude of the difference, that is, with the difference regardless of sign.

However, in many experiments, our primary concern may be with the direction of the difference rather than with its existence in absolute terms. For example, if we plan an experiment to study the effect of coaching work on computational skill in mathematics, we take two groups—the experimental group, which is provided an extra one hour coaching work in mathematics, and the control group, which is not provided with such a drill. Here, we have reason to believe that the experimental group will score higher on the mathematical computation ability test which is given at the end of the session.

In our experiment we are interested in finding out the gain in the acquisition of mathematical computation skill (we are not interested in the loss, as it seldom happens that coaching will decrease the level of computation skill).

In cases like these, we make use of the one-tailed or directional test, rather than the two-tailed or non-directional test to test the significance of difference between means.

Consequently, the meaning of null hypothesis, restricted to an hypothesis of no difference with two-tailed test, will be somewhat extended in a one-tailed test to include the direction-positive or negative-of the difference between means.

Self Assessment Questions

1) How is size of sample important in setting up level of confidence?

   .....

   .....

2) What is one tailed tests of significance? Explain with examples.

   .....

   .....

3) What is a two tailed test? When is it useful? Give suitable examples.

   .....

   .....

4.7 STEPS IN SETTING UP THE LEVEL OF SIGNIFICANCE

i) State the null hypothesis and the alternative hypothesis. (Note: The goal of inferential statistics is to make general statements about the population by using sample data. Therefore in testing hypothesis, we make our predictions about the population parameters).

ii) Set the criteria for a decision.

iii) Level of significance or alpha level for the hypothesis test: This is represented by μ which is the probability used to define the very unlikely sample outcomes, if the null hypothesis is true.

In hypothesis testing, the set of potential samples is divided into those that are likely to be obtained and those that are very unlikely if the hypothesis is true.

iv) Critical Region: This is the region which is composed of extreme sample values that are very unlikely outcomes if the null hypothesis is true. The boundaries for

the critical region are determined by the alpha level. If sample data fall in the critical region, the null hypothesis is rejected. The cut off level that is set affects the outcome of the research.

v) Collect data and compute sample statistics using the formula given below

$$z = \frac{\overline{x} - \mu}{\sigma_x}$$

where, $\overline{x}$ = sample mean,

μ = hypothesised population mean, and

$\acute{o}_x$ = standard error between $\overline{x}$ and μ.

vi) Make a decision and write down the decision rule.

Z-Score is called a test statistics. The purpose of a test statistics is to determine whether the result of a research study (the obtained difference) is more than what would be expected by the chance alone.

$$z = \frac{\text{Obtained difference}}{\text{Difference due to chance}}$$

Now suppose a manufacturer, produces some type of articles of good quality. A purchaser by chance selects a sample randomly. It so happens that the sample contains many defective articles and it leads the purchaser to reject the whole product. Now, the manufacturer suffers a loss even though he has produced a good article of quality. Therefore, this Type-I error is called "*producers risk*".

On the other hand, if we accept the entire lot on the basis of a sample and the lot is not really good, the consumers are put to loss. Therefore, this Type-Il error is called the "*consumers risk*".

In practical situations, still other aspects are considered while accepting or rejecting a lot. The risks involved for both producer and consumer are compared. Then Type-I andType-Il errors are fixed; and a decision is reached.

In summary, the following procedure is recommended for formulating hypotheses and stating conclusions.

4.7.1 Formulating Hypothesis and Stating Conclusions

i) State the hypothesis as the alternative hypothesis H_1.

ii) The null hypothesis, Ho, will be the opposite of H_1 and will contain an equality sign.

iii) If the sample evidence supports the alternative hypothesis, the null hypothesis will be rejected and the probability of having made an incorrect decision (when in fact Ho is true) is α, a quantity that can be manipulated to be as small as the researcher wishes.

iv) If the sample does not provide sufficient evidence to support the alternative hypothesis, then conclude that the null hypothesis cannot be rejected on the basis of your sample. In this situation, you may wish to collect more information about the phenomenon under study.

An example is given below:

Example

The logic used in hypothesis testing has often been likened to that used in the courtroom in which a defendant is on trial for committing a crime.

i) Formulate appropriate null and alternative hypotheses for judging the guilt or innocence of the defendant.

ii) Interpret the Type-I and Type-Il errors in this context.

iii) If you were the defendant, would you want á to be small or large? Explain.

Solution

i) Under a judicial system, a defendant is "innocent until proven guilty". That is, the burden of proof is not on the defendant to prove his or her innocence; rather, the court must collect sufficient evidence to support the claim that the defendant is guilty. Thus, the null and alternative hypotheses would be

H_0 : Defendant is innocent

H_1: Defendant is guilty

ii) The four possible outcomes are shown in the table below. A Type-I error would be to conclude that the defendant is guilty, when in fact he or she is innocent; a Type-II error would be to conclude that the defendant is innocent, when in fact he or she is guilty.

Table : Conclusions and Consequences

		Decision of Court	
		Defendant is Innocent	Defendant is Guilty
True State of Nature	Defendant if Innocent Defendant is Guilty	*Correct decision* Type-I error	*Type-II error* Correct decision

iii) Most people would probably agree that the Type-I error in this situation is by far the more serious. Thus, we would want α, the probability of committing a Type-I error, to be very small indeed.

A convention that is generally observed when formulating the null and alternative hypotheses of any statistical test is to state Ho so that the possible error of incorrectly rejecting Ho (Type-I error) is considered more serious than the possible error of incorrectly failing to reject Ho (Type-II error).

In many cases, the decision as to which type of error is more serious is admittedly not as clear-cut though experience will help to minimize this potential difficulty.

4.7.2 Types of Errors for a Hypothesis Test

The goal of any hypothesis testing is to make a decision. In particular, we will decide whether to reject the null hypothesis, Ho, in favour of the alternative hypothesis, H_1. Although we would like always to be able to make a correct decision, we must remember that the decision will be based on sample information, and thus we are subject to make one of two types of error, as defined in table below.

The null hypothesis can be either true or false. Further, we will make a conclusion

either to reject or not to reject the null hypothesis. Thus, there are four possible situations that may arise in testing a hypothesis as shown in table .

Table: Conclusions and Consequences for Testing a Hypothesis

		Decision of Court	
		Defendant is Innocent	Defendant is Guilty
True "State of Nature"	Null Hypothesis	Correct Conclusion	*Type-I error*
	Alternative Hypothesis	*Type-II error*	Correct Conclusion

The kind of error that can be made depends on the actual state of affairs (which, of course, is unknown to the investigator). Note that we risk a Type-I error only if the null hypothesis is rejected, and we risk a Type-II error only if the null hypothesis is not rejected.

Thus, we may make no error, or we may make either a Type-I error (with probability a), or a Type-II error (with probability 3), but not both. We don't know which type of error corresponds to actuality and so would like to keep the probabilities of both types of errors small.

Remember that as α increases, β decreases, similarly, as β increases, α decreases. The only way to reduce α and β simultaneously is to increase the amount of information available in the sample, i.e. to increase the sample size.

You may note that we have carefully avoided stating a decision in terms of "accept the null hypothesis H_0". Instead, if the sample does not provide enough evidence to support the alternative hypothesis H_1 we prefer a decision "not to reject H_0".

This is because, if we were to "accept H_0", the reliability of the conclusion would be measured by α, the probability of Type-II error. However, the value of β is not constant, but depends on the specific alternative value of the parameter and is difficult to compute in most testing situations.

Self Assessment Questions

1) Elucidate the steps in setting up the level of significance.

   .....

   .....

2) How do you formulate hypothesis and state the conclusions?

   .....

   .....

3) Explain the concept if α increases β decreases. If β increases α decreases.

   .....

   .....

4) Why β is more important than β? Explain.

   .....

   .....

4.8 LET US SUM UP

In this unit, we pointed out how drawing conclusions about a population on the basis of sample information is called statistical inference. Here we have basically two things to do: statistical estimation and hypothesis testing.

An estimate of an unknown parameter could be either a point or an interval. Sample mean is usually taken as a point estimate of population mean. On the other hand, in interval estimation we construct two limits (upper and lower) around the sample mean. We can say with stipulated level of confidence that the population mean, which we do not know; is likely to remain within the confidence interval.

We learnt about confidence interval and how to set the same. In order to construct confidence interval we need to know the population variance or its estimate. When we know population variance, we apply normal distribution to construct the confidence interval. In cases where population variance is not known, we use t distribution for the above purpose.

Remember that when sample size is large (n>30) t-distribution approximates normal distribution. Thus for large samples, even if population variance is not known, we can use normal distribution for estimation of confidence interval on the basis of sample mean and sample variance.

Subsequently we discussed the methods of testing a hypothesis and drawing conclusions about the population. Hypothesis is a simple statement (assertion or claim) about the va1ue assumed by the parameter. We test a hypothesis on the basis of sample information available to us. In this Unit we considered two situations: i) description of a single sample, and ii) comparison between two samples.

In the case of qualitative data we pointed out how we cannot have parametric values and hypothesis testing on the basis of z statistic or t-statistic cannot be performed.

4.9 UNIT END QUESTIONS

1) What do you mean by a null hypothesis?

2) What is significance of size of sample is hypothesis testing?

3) Write down two levels of significance which are mainly used in hypothesis testing?

4) Write down a short note on level of significance.

4.10 GLOSSARY

Contingency Table : A two-way table to present bivariate data. It is called contingency table because we try to find whether one variable is contingent upon the other variable.

Degrees of Freedom : It refers to the number of pieces of independent information that are required to compute some characteristic of a given set of observations.

Estimation : It is the method of prediction about parameter values on the basis of sample statistics.

Expected Frequency : It is the expected cell frequency under the assumption that both the variables are independent.

Nominal Variable : Such a variable takes qualitative values and do not have any ordering relationships among them. For example, gender is a nominal variable taking only the qualitative values, male and female; there is no ordering in 'male' and 'female' status. A nominal variable is also called an attribute.

Parameter : It is a measure of some characteristic of the population.

Population : It is the entire collection of units of a specified type in a given place and at a particular point of time.

Random Sampling : It is a procedure where every member of the population has a definite chance or probability of being selected in the sample. It is also called probability sampling. Random sampling could be of many types: simple random sampling, systematic random sampling and stratified random sampling.

Sample : It is a sub-set of the population. It can be drawn from the population in a scientific manner by applying the rules of probability so that personal bias is eliminated. Many samples can be drawn from a population and there are many methods of drawing a sample.

Sampling Distribution : It is the relative frequency or probability distribution of the values of a statistic when the number of samples tends to infinity.

Sampling Error : In the sampling method, we try to approximate some feature of a given population from a sample drawn from it. Now, since in the sample all the members of the population are not included, howsoever close the approximation is, it is not identical to the required population feature and some error is committed. This error is called the sampling error.

Significance Level : There may be certain samples where population mean would not remain within the confidence interval around sample mean. The percentage (probability) of such cases is called significance level. It is usually denoted by.

4.11 SUGGESTED READINGS

Asthana H.S, and Bhushan. B. (2007) *Statistics for Social Sciences* (with SPSS Applications).

B.L. Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Guilford, J.P. (1965); *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill Book Company.

Gupta, S.C. (1990) *Fundamentals of Statistics*.

Siegel, S. (1956): *Non-parametric Statistics for Behavioural Sciences*. Tokyo: McGraw Hill Hoga Kunsa Ltd.

UNIT 1 PRODUCT MOMENT COEFFICIENT OF CORRELATION

Structure

1.0 Introduction

1.1 Objectives

1.2 Correlation: Meaning and Interpretation

- 1.2.1 Scatter Diagram: Graphical Presentation of Relationship
- 1.2.2 Correlation: Linear and Non-Linear Relationship
- 1.2.3 Direction of Correlation: Positive and Negative
- 1.2.4 Correlation: The Strength of Relationship
- 1.2.5 Measurements of Correlation
- 1.2.6 Correlation and Causality

1.3 Pearson's Product Moment Coefficient of Correlation

- 1.3.1 Variance and Covariance: Building Blocks of Correlations
- 1.3.2 Equations for Pearson's Product Moment Coefficient of Correlation
- 1.3.3 Numerical Example
- 1.3.4 Significance Testing of Pearson's Correlation Coefficient
- 1.3.5 Adjusted *r*
- 1.3.6 Assumptions for Significance Testing
- 1.3.7 Ramifications in the Interpretation of Pearson's r
- 1.3.8 Restricted Range

1.4 Unreliability of Measurement

- 1.4.1 Outliers
- 1.4.2 Curvilinearity

1.5 Using Raw Score Method for Calculating r

- 1.5.1 Formulas for Raw Score
- 1.5.2 Solved Numerical for Raw Score Formula

1.6 Let Us Sum Up

1.7 Unit End Questions

1.8 Suggested Readings

1.0 INTRODUCTION

We measure psychological attributes of people by using tests and scales in order to describe individuals. There are times when you realise that increment in one of the characteristics is associated with increment in other characteristic as well. For example, individuals who are more optimistic about the future are more likely to be happy. On the other hand, those who are less optimistic about future (i.e., pessimistic about it) are less likely to be happy. You would realise that as one variable is increasing, the other is also increasing and as the one is decreasing the other is also decreasing. In the statistical language it is referred to as correlation. It is a description of "relationship" or "association" between two variables (more than two variables can also be correlated, we will see it in multiple correlation).

In this unit you will be learning about direction of Correlation that is, Positive and Negative and zero correlation. You will also learn about the strength of correlation and how to measure correlation. Specifically you will be learning Pearson's Product Moment Coefficient of Correlation and how to interpret this correlation coefficient. You will also learn about the ramifications of the Pearson's r. You will also learn the coefficient of correlation equations with numerical examples.

1.1 OBJECTIVES

After reading and doing exercises in this unit, you will be able to:

- describe and explain concept of correlation;
- plot the scatter diagram;
- explain the concept of direction, and strength of relationship;
- differentiate between various measures of correlations;
- analyse conceptual issues in correlation and causality;
- describe problems suitable for correlation analysis;
- describe and explain concept of Pearson's Product Moment Correlation;
- compute and interpret Pearson's correlation by deviation score method and raw score method; and
- test the significance and apply the correlation to the real data.

1.2 CORRELATION: MEANING AND INTERPRETATION

Correlation is a measure of association between to variables. Typically, one variable is denoted as X and the other variable is denoted as Y. The relationship between these variables is assessed by correlation coefficient. Look at the earlier example of optimism and happiness. It states the relationship between one variable, optimism (X) and other variable, happiness (Y). Similarly, following statements are example of correlations:

As the *intelligence* (IQ) increases the *marks* obtained increases.

As the *introversion* increases *number of friends* decreases.

More the *anxiety* a person experiences, weaker the *adjustment* with the stress.

As the score on *openness to experience* increases, scores on *creativity* test also increase.

More the *income*, more the *expenditure*.

On a reasoning task, as the *accuracy* increases, the *speed* decreases.

As the *cost* increases the *sales* decrease.

Those who are good at *mathematics* are likely to be good at *science*.

As the age of the child increases, the *problems solving capacity* increase.

More the *practice*, better the *performance*.

All the above statements exemplify the correlation between two variables. The variables are shown in *italics*. In this first section, we shall introduce ourselves to the concept of correlation.

1.2.1 Scatter Diagram: Graphical Presentation of Relationship

Scatter diagram (also called as *scatterplot, scattergram,* or *scatter*) is one way to study the relationship between two variables. Scatter diagram is to plot pairs of values of subjects (observations) on a graph. Let's look at the following data of five subject, A to E (Table 1.1). Their scores on intelligence and scores on reasoning task are provided. The same data is used to plot a scatter diagram shown in Figure 1.1. Now, I shall quickly explain 'how to draw the scatter diagram'.

Table 1: Data of five subjects on intelligence and scores of reasoning

Subject	Intelligence	Scores on reasoning task
A	104	12
B	127	25
C	109	18
D	135	31
E	116	19

Step 1. Plotting the Axes

Draw the x and y axis on the graph and plot one variable on x-axis and another on y-axis.

(Although, correlation analyses do not restrict you from plotting any variable on any axis, plot the causal variable on x-axis in case of implicitly assumed cause-effect relationship.)

Also note that correlation does not necessarily imply causality.

Step 2. Range of Values

Decide the range of values depending on your data.

Begin from higher or lower value than zero.

Conventionally, the scatterplot is square.

So plot x and y values about the same length.

Step 3. Identify the pairs of values

Identify the pairs of values.

A pair of value is obtained from a data.

A pair of values is created by taking a one value on first variable and corresponding value on second variable.

Step 4. Plotting the graph

Now, locate these pairs in the graph.

Find an intersection point of x and y in the graph for each pair.

Mark it by a clear dot (or any symbol you like for example, star).

Then take second pair and so on.

The scatterplot shown below is based on the data given in table 1. (Refer to Figure 1).

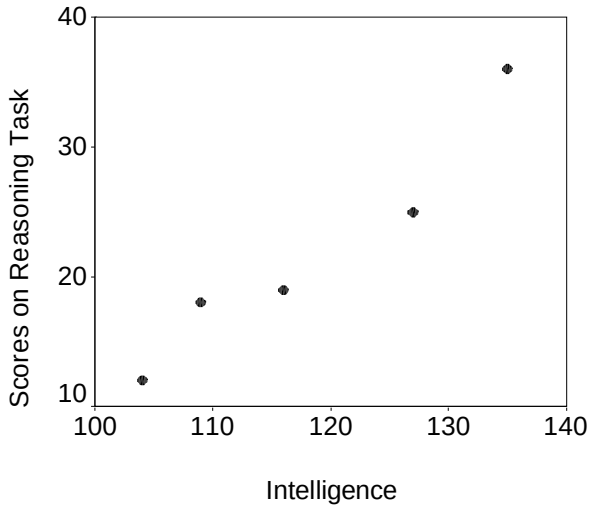


Fig. 1: Scatter diagram depicting relationship between intelligence and score on reasoning task

The graph shown above is scatterplot representing the relationship between intelligence and the scores on reasoning task. We have plotted intelligence on x-axis because it is a cause of the performance on the reasoning task. The scores on reasoning have started from 100 instead of zero simply because the smallest score on intelligence is 104 which is far away from zero. We have also started the range of reasoning scores from 10 since the lowest score on reasoning is 12. Then we have plotted the pair of scores. For example, subject A has score of 104 on intelligence and 12 on reasoning so we get x,y pair of 104,12. We have plotted this pair on the point of intersection between these two scores in the graph by a dot. This is the lowest dot at the left side of the graph. You can try to practice the scatter by using the data given in the practice.

1.2.2 Correlation: Linear and Non-Linear Relationship

The relationship between two variables can be of various types. Broadly, they can be classified as linear and nonlinear relationships. In this section we shall try to understand the linear and nonlinear relationships.

Linear Relationship

One of the basic forms of relationship is linear relationship. *Linear* relationship can be expressed as a relationship between two variables that can be plotted as a *straight* line. The linear relationship can be expressed in the following equation (eq. 1.1):

$$Y = á + â\ X \qquad \text{(eq. 1.1)}$$

In the equation 1.1,

- Y is a dependent variable (variable on y-axis),

- á (alpha) is a constant or Y intercept of straight line,
- â (beta) is slope of the line and
- X is independent variable (variable on x-axis).

We again plot scatter with the line that best fits for the data shown in table 1. So you can understand the linearity of the relationship. Figure 2 shows the scatter of the same data. In addition, it shows the line which is best fit line for the data. This line is plotted by using the method of least squares. We will learn more about it later (Unit 4). Figure 2 shows that there is a linear relationship between two variables, intelligence and Scores on Reasoning Task. The graph also shows the straight line relationship indicating linear relation.

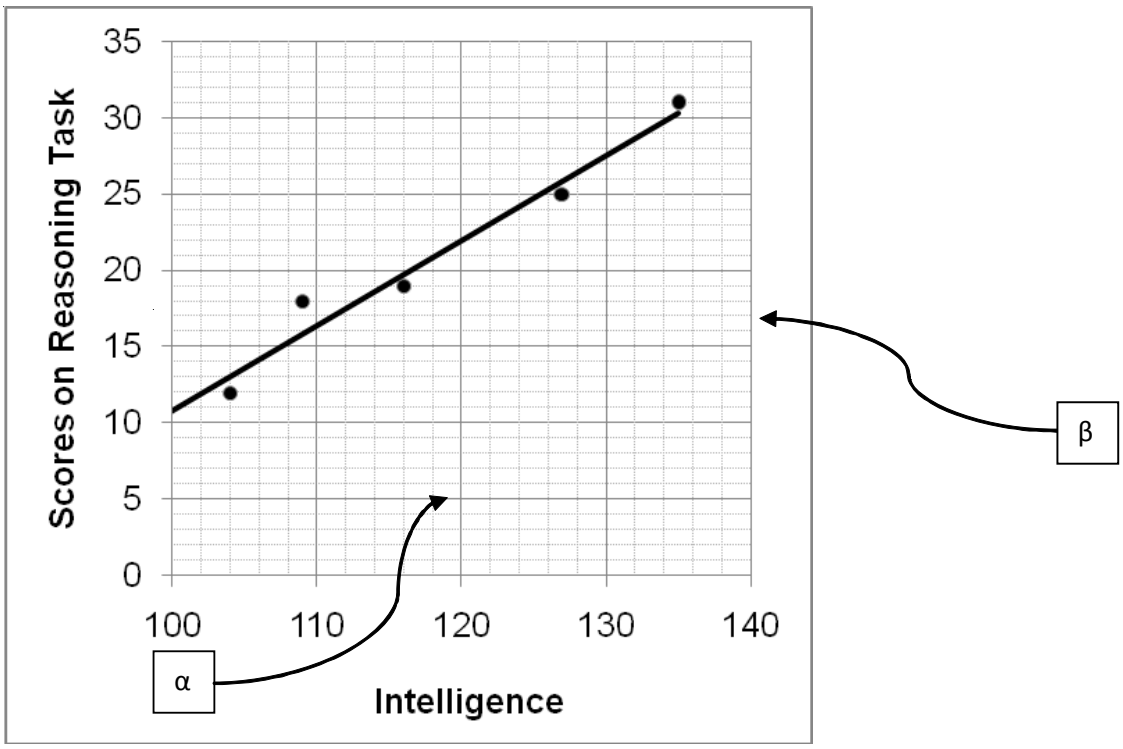


Fig. 2: Scatter showing linearity of the relationship between Intelligence and Scores on Reasoning Task

Non-linear Relationship

There are other forms of relationships as well. They are called as curvilinear or non-linear relationships. The Yorkes-Dodson Law, Steven's Power Law in Psychophysics, etc. are good examples of non-linear relationships. The relationship between stress and performance is popularly known as Yorkes-Dodson Law. It suggests that the performance is poor when the stress is too little or too much. It improves when the stress is moderate. Figure 3 shows this relationship. The *non-linear* relationships, cannot be plotted as a *straight line*.

The performance is poor at extremes and improves with moderate stress. This is one type of curvilinear relationship.

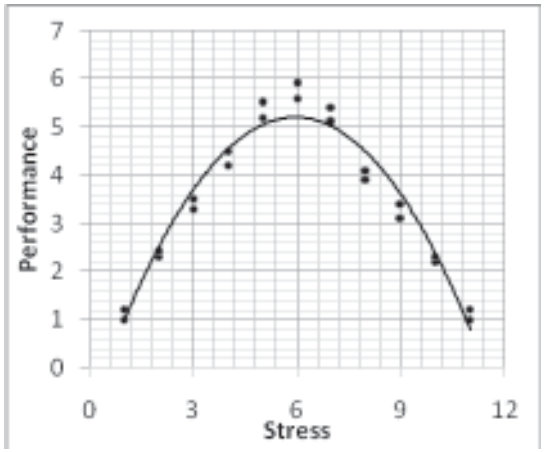


Fig. 3: Typical relationship between stress and performance

The curvilinear relationships are of various types (cubic, quadratic, polynomial, exponential, etc.). The point we need to note is that *relationships can be of various types*. This block discussed only *linear* relationships. Other forms of relationship are *not* discussed. The types of correlation presented in this block represent linear relationships. Pearson's product-moment correlation, Spearman's rho, etc. are linear correlations.

The Stevens' Power Law states that $r = cs^b$ where, r is sensation, s is stimulus, c and b are constants and coefficients, respectively. This is obviously a non-linear relationship between stimulus and sensation. Although, a reader who can recall some basic mathematics of 10^{th} grade can easily understand that by taking the log of both sides, the equation can be converted into linear equation.

1.2.3 Direction of Correlation: Positive and Negative

The direction of the relationship is an important aspect of the description of relationship. If the two variables are correlated then the relationship is either positive or negative. The absence of relationship indicates "zero correlation". Let's look at the positive, negative and zero correlation.

Positive Correlation

The positive correlation indicates that as the values of one variable increases the values of other variable also increase. Consequently, as the values of one variable decreases, the values of other variable also decrease. This means that both the variables move in the samer direction. For example,

a) As the *intelligence* (IQ) increases the *marks* obtained increases.

b) As *income* increases, the *expenditure* increases.

The figure 4 shows *scatterplot* of the positive relationship. You will realise that the higher scores on X axis are associated with higher score on Y axis and lower scores on X axis are generally associated with lower score on Y axis. In the 'a' example, higher scores on *intelligence* are associated with the higher score on *marks obtained*. Similarly, as the scores on *intelligence* drops down, the *marks obtained* has also dropped down.

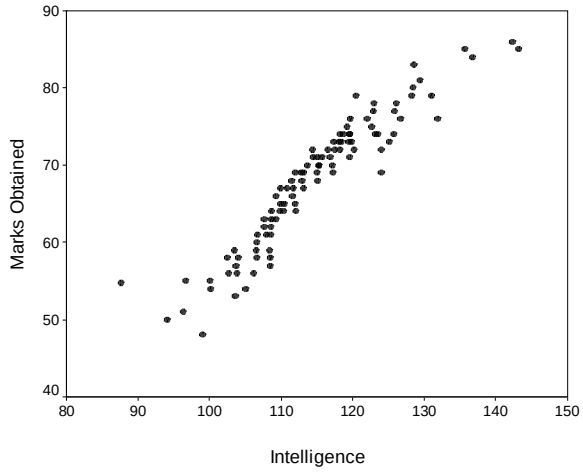


Fig. 4: Positive correlation: Scatter showing the positive correlation between *intelligence* and *marks obtained.*

Negative Correlation

The Negative correlation indicates that as the values of one variable increases, the values of the other variable decrease. Consequently, as the values of one variable decreases, the values of the other variable increase. This means that two variables move in the opposite direction. For example,

a) As the *intelligence* (IQ) increases the *errors on reasoning task* decreases.

b) As *hope* increases, *depression* decreases.

Figure 5 shows *scatterplot* of the negative relationship. You will realise that the higher scores on X axis are associated with lower scores on Y axis and lower scores on X axis are generally associated with higher score on Y axis.

In the 'a' example, higher scores on *intelligence* are associated with the lower score on *errors on reasoning task*. Similarly, as the scores on *intelligence* drops down, the *errors on reasoning task* have gone up.

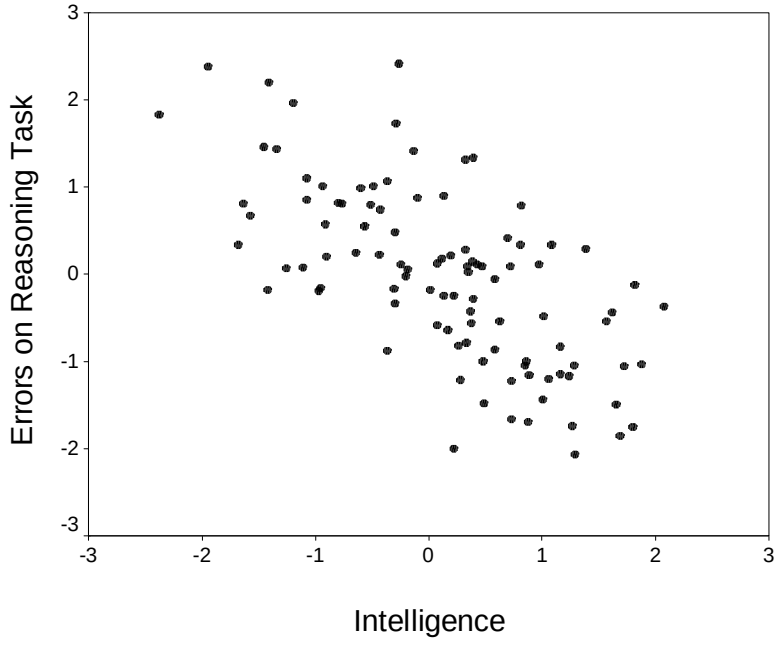


Fig. 5: Negative correlation: Scatter showing the negative correlation between *intelligence* and *errors on reasoning task*

No Relationship

Until now you have learned about the positive and negative correlations. Apart from positive and negative correlations, it is also possible that there is no relationship between *x* and *y*. That is the two variables do not share any relationship. If they do not share any relationship (that is, technically the correlation coefficient is zero), then, obviously, the direction of the correlation is neither positive nor negative. It is often called as zero correlation or no correlation.

(Please note that 'zero order correlation' is a different term than 'zero correlation' which we will discuss afterwards).

For example, guess the relationship between shoe size and intelligence?

This sounds an erratic question because there is no reason for any relationship between them. So there is no relationship between these two variables.

The data of one hundred individuals is plotted in Figure 6. It shows the scatterplot for no relationship.

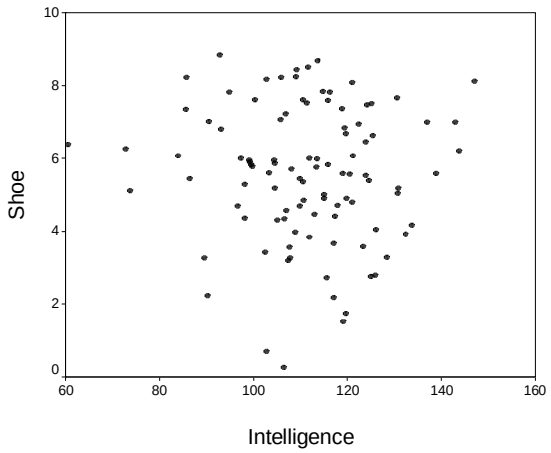


Fig. 6: Scatter between shoe size and intelligence of the individual

1.2.4 Correlation: The Strength of Relationship

You have so far learnt the direction of relationship between two variables. Any curious reader will ask a question "how strong is the relationship between the two variables?" For example, if you correlate intelligence with scores on reasoning, and creativity, what kind of relationship will you expect?

Obviously, the relationship between intelligence and reasoning as well as the relationship between intelligence and creativity are positive. At the same time the correlation coefficient (described in the following section) is higher for intelligence and reasoning than for intelligence and creativity, and therefore we realise that the relationship between intelligence and reasoning is stronger than relationship between intelligence and creativity. The strength of relationship between the two variables is an important information to interpret the relationship.

Correlation Coefficient

The correlation between any two variables is expressed in terms of a number, usually called as correlation coefficient. The correlation coefficient is denoted by various symbols depending on the type of correlation. The most common is '*r*' (small 'r') indicating the Pearson's product-moment correlation coefficient.

The representation of correlation between X and Y is r_{xy}.

The range of the correlation coefficient is from –1.00 to + 1.00.

It may take any value between these numbers including, for example, – 0.72, – 0.61, – 0.35, + 0.02, + 0.31, + 0.98, etc.

If the correlation coefficient is 1, then relationship between the two variables is perfect.

This will happen if the correlation coefficient is – 1 or + 1.

As the correlation coefficient moves nearer to + 1 or – 1, the strength of relationship between the two variables increases.

If the correlation coefficient moves away from the + 1 or – 1, then the strength of relationship between two variables decreases (that is, it becomes weak).

So correlation coefficient of + 0.87 (and similarly – 0.82, – 0.87, etc.) shows strong association between the two variables. Whereas, correlation coefficient of + 0.24 or – 0.24 will indicate weak relationship. Figure 7 indicates the range of correlation coefficient.

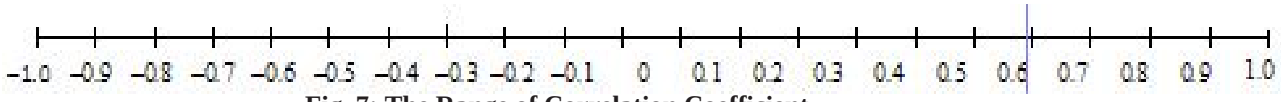


Fig. 7: The Range of Correlation Coefficient.

You can understand the strength of association as the common variance between two correlated variables. The correlation coefficient is NOT percentage.

So correlation of 0.30 does NOT mean it is 30% variance.

The shared variance between two correlated variables can be calculated. Let me explain this point. See, every variable has variance. We denote it as S_x^2 (variance of X). Similarly, Y also has its own variance (S_y^2). In the previous block you have learned to compute them. From the complete variance of X, it shares some variance with Y. It is called covariance.

The Figure 8 shown below explains the concept of shared variance. The circle X indicates the variance of X. Similarly, the circle Y indicates the variance of Y. The overlapping part of X and Y, indicated by shaded lines, shows the shared variance between X and Y. One can compute the shared variance.

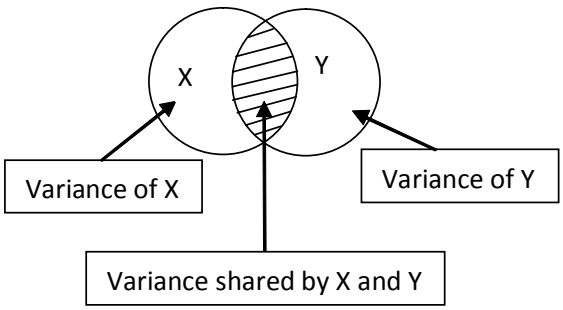


Fig. 8: Covariance indicates the degree to which X shares variance with Y

To calculate the percentage of shared variance between X and Y (common variance), one needs to square the correlation coefficient (*r*). The formula is given below:

Percentage of common variance between X and Y = $r_{xy}^2 \times 100$ (eq. 1.2)

For instance, if the correlation between X and Y is 0.50 then the percent of variation shared by X and Y can be calculated by using equation 1.2 as follows.

Percentage of common variance between X and Y = $r_{xy}^2 \times 100 = 0.50^2 \times 100 = 0.25 \times 100 = 25\%$

It indicates that, if the correlation between X and Y is 0.50 then 25% of the variance is shared by the two variables, X and Y. You would note that this formula is applicable to negative correlations as well. For instance, if $r_{xy} = -0.81$, then shared variance is:

Percentage of common variance between X and Y = $\times 100 = -0.81^2 \times 100 = 0.6561 \times 100 = 65.61\%$

1.2.5 Measurements of Correlation

Correlation coefficient can be calculated by various ways. The correlation coefficient is a description of association between two variables in the sample. So it is a descriptive statistics. Various ways to compute correlation simply indicate the degree of association between variables *in the sample*. The distributional assumptions are *not* required to compute correlation as a descriptive statistics. So it is not a parametric or nonparametric statistics.

The calculated sample correlation coefficient can be used to estimate population correlation coefficient.

The sample correlation coefficient is usually denoted by symbol '*r*'.

The population correlation coefficient is denoted by symbol '$\tilde{n}$'.

It is Greek letter *rho($\tilde{n}$)*, pronounced as *row* (Spearman's correlation coefficient is also symbolised as *rho*.

This may create some confusion among the readers. Therefore, I shall use symbol r_s for Spearman's *rho* as a sample statistics and $\tilde{n}_s$ to indicate the population value of the Spearman's *rho*.

Henceforth, I shall also clearly mention the meaning with which $\tilde{n}$ is used in this block.

- When the population correlation coefficient is estimated from sample correlation coefficient.
- then the correlation coefficient becomes an inferential statistic.
- Inference about population correlation ($\tilde{n}$) is drawn from sample statistics (*r*).
- The population correlation ($\tilde{n}$) is always unknown.
- What is known is sample correlation (*r*).
- The population indices are called as parameters and the sample indices are called as statistics.
- So $\tilde{n}$ is a parameter and *r* is a statistics.

While inferring a parameter from sample, certain distributional assumptions are required. From this, you can understand that the descriptive use of the correlation coefficient does not require any distributional assumptions.

The most popular way to compute correlation is 'Pearson's Product Moment Correlation (*r*)'. This correlation coefficient can be computed when the data on both the variables is on at least equal interval scale or ratio scale.

Apart from Pearson's correlation there are various other ways to compute correlation. Spearman's Rank Order Correlation or Spearman's *rho* (r_s) is useful correlation coefficient when the data is in rank order.

Similarly, Kendall's *tau* (ô) is a useful correlation coefficient for rank-order data.

Biserial, Point Biserial, Tetrachoric, and Phi coefficient, are the correlations that are useful under special circumstances.

Apart from these, multiple correlations, part correlation and partial correlation are useful ways to understand the associations (Please note that the last three require more than two variables).

1.2.6 Correlation and Causality

The correlation does not necessarily imply causality. But, if the correlation between two variables is high then it might indicate the causality. If X and Y are correlated, then there are three different ways in which the relationship between two variables can be understood in terms of causality.

1) X is a cause of Y.

2) Y is cause of X.

3) Both, X and Y are caused by another variable Z.

However, causality can be inferred from the correlations.

Regression analysis, path analysis, structural equation modeling, are some examples where correlations are employed in order to understand causality.

1.3 PEARSON'S PRODUCT MOMENT COEFFICIENT OF CORRELATION

The Person's correlation coefficient was developed by Karl Pearson in 1886. Person was a editor of "Biometrica" which is a leading journal in statistics. Pearson was a close associate of psychologist Sir Francis Galton. The Pearson's correlation coefficient is usually calculated for two continuous variables. If either or both the variables are not continuous, then other statistical procedures are to be used. Some of them are equivalent to Pearson's correlation and others are not. We shall learn about these procedure after learning the Pearson's Correlation coefficient.

1.3.1 Variance and Covariance: Building Blocks of Correlations

Understanding product moment correlation coefficient requires understanding of mean, variance and covariance. We shall understand them once again in order to understand correlation.

Mean : Mean of variable X (symbolised as $\overline{X}$) is sum of scores ($\sum_{i=1}^{n} X_i$) divided by number of observations (n). The mean is calculated in following way.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \text{(eq. 1.3)}$$

You have learned this in the first block. We will need to use this as a basic element to compute correlation.

Variance

The variance of a variable X (symbolised as S_X^2) is the sum of squares of the deviations of each X score from the mean of X ($\sum (X - \overline{X})^2$)divided by number of observations (*n*).

$$S_x^2 = \frac{\sum (X - \overline{X})^2}{n} \qquad \text{(eq. 1.4)}$$

You have already learned that standard deviation of variable X , symbolised as , is square root of variance of X, symbolised as .

Covariance

The covariance between X and Y (or S_{XY}) can be stated as

$$Cov_{XY} = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{n} \qquad \text{(eq. 1.5)}$$

Covariance is a number that indicates the association between two variables. To compute covariance, deviation of each score on X from its mean ( ) and deviation of each score on Y from its mean ($\overline{y}$) is initially calculated.

Then products of these deviations are obtained.

Then, these products are summated.

This sum gives us the numerator for covariance.

Divide this sum by number of observations (*n*). The resulting number is covariance.

1.3.2 Equations for Pearson's Product Moment Coefficient of Correlation

Having revised the concepts, we shall now learn to compute the Pearson's Correlation Coefficient.

Formula

Since we have already learned to compute the covariance, the simplest way to define Pearson's correlation is…

$$r = \frac{Cov_{XY}}{S_X S_Y} \qquad \text{(eq. 1.6)}$$

Where,

the Cov_{XY} is covariance between X and Y,

S_X is standard deviation of X

S_Y is standard deviation of Y.

Since, it can be shown that Cov_{XY} is always smaller than or equal to $S_X S_Y$, the maximum value of correlation coefficient is bound to be 1.

The sign of Pearson's *r* depends on the sign of Cov_{XY}.

If the Cov_{XY} is negative, then *r* will be negative and

if Cov_{XY} is positive then *r* will be a positive value.

The denominator of this formula ($S_X S_Y$) is always positive. This is the reason for a – 1 to + 1 range of correlation coefficient. By substituting covariance equation (eq. 1.5) for covariance we can rewrite equation 1.6 as

$$r = \frac{\frac{\sum (X - \bar{X})(Y - \bar{Y})}{n}}{S_X S_Y} \qquad \text{(eq. 1.7)}$$

By following a simple rule, a÷ b ÷ c = a ÷ (b × c), we can rewrite equation 1.7 as follows.

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n S_X S_Y} \qquad \text{(eq. 1.8)}$$

1.3.3 Numerical Example

Now we shall use this formula to compute Pearson's correlation coefficient. For this purpose we will use the following data. The cognitive theory of depression argues that hopelessness is associated with depression. Aron Beck developed instruments to measure depression and hopelessness. The BHS (Beck Hopelessness Scale) and the BDI (Beck Depression Inventory) are measures of hopelessness and depression, respectively.

Let's take a hypothetical data of 10 individuals on whom these scales were administered. (In reality, such a small data is not sufficient to make sense of correlation; roughly, at least a data of 50 to 100 observations is required). We can hypothesize that the correlation between hopelessness and depression will be positive. This hypothetical data is given below in table 2.

Table 2: Hypothetical data of 10 subjects on BHS and BDI

Subject	BHS (X)	BDI (Y)	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
1	11	13	0	1	0	1	0
2	13	16	2	4	4	16	8
3	16	14	5	2	25	4	10
4	9	10	-2	-2	4	4	4
5	6	8	-5	-4	25	16	20
6	17	16	6	4	36	16	24
7	7	9	-4	-3	16	9	12
8	12	12	1	0	1	0	0
9	5	7	-6	-5	36	25	30
10	14	15	3	3	9	9	9
n = 10	$\sum X$ =110	$\sum Y$ =120			$\sum(X-\bar{X})^2$ = 156	$\sum(Y-\bar{Y})^2$ = 100	$\sum(X-\bar{X})(Y-\bar{Y})$ = 117
	$\bar{X}=11$	$\bar{Y}=12$					

$$S_X = \sqrt{\sum(X-\bar{X})^2/n} = 4.16$$

$$S_Y = \sqrt{\sum(Y-\bar{Y})^2/n} = 3.33$$

$$r = \sum(X-\bar{X})(Y-\bar{Y})/nS_XS_Y = 117/(10)(4.16)(3.33) = +0.937$$

Step 1. You need scores of subjects on two variables. We have scores on ten subjects on two variables, BHS and BDI.

Step 2. Then list the pairs of scores on two variables in two columns. The order will not make any difference. Remember, same individuals' two scores should be kept together. Label one variable as X and other as Y. We label BHS as X and BDI as Y.

Step 3. Compute the mean of variable X and variable Y. It was found to be 11 and 12 respectively.

Step 4. Compute the deviation of each X score from its mean ( ) and each Y score from its own mean ($\bar{Y}$). This is shown in the column labeled as $X-\bar{X}$ and $Y-\bar{Y}$. As you have learned earlier, the sum of these columns has to be zero.

Step 5. Compute the square of $x-\bar{x}$ and $y-\bar{y}$.

This is shown in next two columns labelled as $(x-\bar{x})^2$ and $(y-\bar{y})^2$.

Step 6. Then compute the sum of these squared deviations of X and Y. The sum of squared deviations for X is 156 and for Y it is 100.

Step 7. Divide them by *n* to obtain the standard deviations for X and Y. The S_x was found to be 4.16. Similarly, the S_y was found to be 3.33.

Step 8. Compute the cross-product of the deviations of X and Y. These cross-products are shown in the last column labeled as $(x - \bar{x})(y - \bar{y})$.

Step 9. Then obtain the sum of these cross-products. It was found to be 117. Now, we have all the elements required for computing *r*.

Step 10. Use the formula of *r* to compute correlation. The sum of the cross-product of deviations is numerator and n, S_x, S_y, are denominators. Compute *r*. the value of *r* is 0.937 in this example.

1.3.4 Significance Testing of Pearson's Correlation Coefficient

Statistical significance testing is testing the hypothesis about the population parameter from sample statistics. When the Pearson's Correlation coefficient is computed as an index of description of relationship between two variables in the sample, the significance testing is not required. The interpretation of correlation from the value and direction is enough.

However, when correlation is computed as an estimate of population correlation, obviously, statistical significance testing is required.

"Whether the obtained sample value of Pearson's correlation coefficient is greater than the value that can be obtained by chance?" is the question answered by statistical significance testing about correlation coefficient.

Different values of correlation can be obtained between any two variables, X and Y, for different samples of different sizes belonging to the same population.

The researcher is not merely interested in knowing the finding in the specific sample on which the data are obtained. But they are interested in estimating the population value of the correlation.

Testing the significance of correlation coefficient is a complex issue. It is because of the distribution of the correlation coefficient. The *t*-distribution and z-distribution are used to test statistical significance of *r*.

The population correlation between X and Y is denoted by $\tilde{n}_{xy}$. The sample correlation is r_{xy}.

As you have learned, we need to write a null hypothesis (H_O) and alternative hypothesis (H_A) for this purpose.

The typical null hypothesis states that population correlation coefficient between X and Y ($\tilde{n}_{xy}$) is zero.

$H_O : \tilde{n}_{xy} = 0$

$H_A : \tilde{n}_{xy} \neq 0$

If we reject the H_O then we accept the alternative (H_A) that the population correlation coefficient is other than zero. It implies that the finding obtained on the data is not a sample-specific error.

Sir Ronald Fisher has developed a method of using *t*-distribution for testing this null hypothesis.

The degrees of freedom (*df*) for this purpose are $n - 2$. Here *n* refers to number of observations.

We can use Appendix C in a statistic book for testing the significance of correlation coefficient. Appendix C provides critical values of correlation coefficients for various degrees of freedom. Let's learn how to use the Appendix C. We shall continue with the example of BHS and BDI.

The correlation between BHS and BDI is +.937 obtained on 10 individuals. We decide to do statistical significance testing at 0.05 level of significance, so our á = .05.

We also decided to apply two-tailed test.

The two-tailed test is used if alternative hypothesis is non-directional, i.e. it does not indicate the direction of correlation coefficient (meaning, it can be positive or negative) and one-tail test is used when alternative is directional (it states that correlation is either positive or negative).

Let us write the null hypothesis and alternative hypothesis:

Null hypothesis

$$H_O : \rho_{\text{BHS BDI}} = 0$$
$$H_A : \rho_{\text{BHS BDI}} \neq 0$$

Now we will calculate the degree of freedom for this example.

$df = n - 2 = 10 - 2 = 8$ (eq. 1.9)

So the *df* for this example are 8. Now look at Appendix C. Look down the leftmost *df* column till you reach *df* = 8. Then look across to find correlation coefficient from column of two-tailed test at level of significance of 0.05. You will reach the critical value of *r*:

$r_{\text{critical}} = 0.632$

Because the obtained (i.e., calculated) correlation value of + 0.937 is greater than critical (i.e., tabled) value, we reject the null hypothesis that there is no correlation between BHS and BDI in the population.

So we accept that there is correlation between BHS and BDI in the population. This method is used regardless of the sign of the correlation coefficient.

We use the absolute value (ignore the sign) of correlation while doing a two-tailed test of significance. The sign is considered while testing one-tailed hypothesis.

For example, if the H_A: $\tilde{n} > 0$, which is a directional hypothesis, then any correlation that is negative will be considered as insignificant.

1.3.5 Adjusted *r*

The Pearson's correlation coefficient (*r*) calculated on the sample is not an unbiased estimate of population coefficient ($\tilde{n}$). When the number of observations (sample size) are small the sample correlation is a biased estimate of population correlation. In order to reduce this bias, the calculated correlation coefficient is adjusted. This is called as adjusted correlation coefficient (r_{adj}).

$$r_{\text{adj}} = \sqrt{1 - \frac{(1 - r^2)(n - 1)}{n - 2}}$$

Where,

r_{adj}= adjusted *r*

r^{2-} = the square of Pearson's Correlation Coefficient obtained on sample,

n = sample size

In case of our data, presented in table 1.2, the correlation between BHS and BDI is +.937 obtained on the sample of 10. The adjusted r can be calculated as follows

$$r_{adj} = \sqrt{1 - \frac{(1 - .937^2)(10 - 1)}{10 - 2}} = \sqrt{1 - \frac{(.1220)(9)}{8}} = \sqrt{1 - 0.1373} = .929$$

The r_{adj} is found to be 0.929. This coefficient is unbiased estimate of population correlation coefficient.

1.3.6 Assumptions for Significance Testing

One may recall that simple descriptive use of correlation coefficient does not involve any assumption about the distribution of either of the variables. However, using correlation as an inferential statistics requires assumptions about X and Y. These assumptions are as follows. Since we are using *t*-distribution, the assumptions would be similar to *t*.

Assumptions:

Independence among the pairs of score

This assumption implies that the scores of any two observations (subjects in case of most of psychological data) are not influenced by each other. Each pair of observation is independent. This is assured when different subjects provides different pairs of observation.

The population of X and the population of Y follow normal distribution and the population pair of scores of X and Y has a normal bivariate distribution.

This assumption states that the population distribution of both the variables (X and Y) is normal. This also means that the pair of scores follows bivariate normal distribution. This assumption can be tested by using statistical tests for normality.

It should be remembered that the *r* is a robust statistics. It implies that some violation of assumption would not influence the distributional properties of *t* and the probability judgments associated with the population correlation.

1.3.7 Ramifications in the Interpretation of Pearson's r

The interpretation of the correlation coefficient depends primarily on two things: direction and strength of relationship. We have already discussed them in detail and hence repetition is avoided.

Direction

If the correlation is positive, then the relationship between two variables is positive. It means that as there is an increase in one there is an increase in another, and as there is a decrement in one there is a decrease in another. When the direction of correlation is negative then the interpretation is vice-versa.

Strength

The strength can be calculated in terms of percentage. We have already learned this formula. So we can convert the correlation coefficient into percentage of common

variance explained and accordingly interpret. For example, if the correlation between X and Y is 0.78, then the common variance shared by X and Y is 60.84 percent.

Usually distinct psychological variables do not share much of the common variance. In fact, the reliability of psychological variables is an issue while interpreting the correlations. Cohen and Cohen have suggested that considering the unreliability of psychological variables, the smaller correlations should also be considered significant.

Although, direction and strength are key pointers while interpreting the correlation, there are finer aspects of interpretation to correlations.

They are :

- range,
- outliers,
- reliability of variables, and
- linearity

The above are some of the important aspects which all obscure the interpretation of correlation coefficient. Let's discuss them one by one.

1.3.8 Restricted Range

It is expected the variables in correlation analysis are measured with full range. For example, suppose we want to study the correlation between hours spent in studies and marks. We are suppose to take students who have varying degree of hours of studies, that is, we need to select students who have spent very little time in studies to the once who have spent great deal of time in studies. Then we will be able to obtain true value of the correlation coefficient.

But suppose we take a very restricted range then the value of the correlation is likely to reduce. Look at the following examples the figure 1.9a and 1.9b.

The figure 1.9a is based on a complete range.

The figure 1.9b is based on the data of students who have studied for longer durations.

The scatter shows that when the range was full, the correlation coefficient was showing positive and high correlation. When the range was restricted, the correlation has reduced drastically.

You can think of some such examples. Suppose, a sports teacher selects 10 students from a group of 100 students on basis of selection criterion, that is their athletic performance.

The actual performance of these ten selected students in the game was correlated with the selection criterion. A very low correlation was obtained between selection criterion and actual game performance. This would naturally mean that the selection criterion is not related with actual game performance. Is it true..? Why so…?

If you look at the edata, you will realise that the range of the scores on selection criterion is extremely restricted (because these ten students were only high scorers) and hence the relationship is weak. So note that whenever you interpret correlations, the range of the variables is large. Otherwise the interpretations will not be valid.

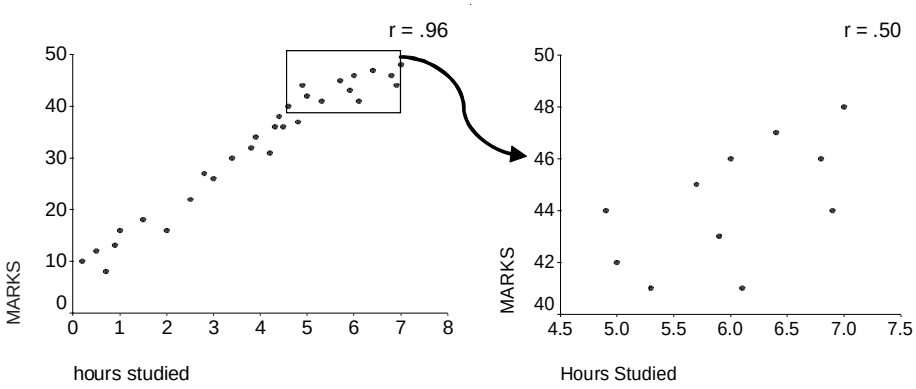


Fig. 1.9a: Scatter showing full range on both variables

Fig. 1.9b: Scatter with restricted range on hours studied

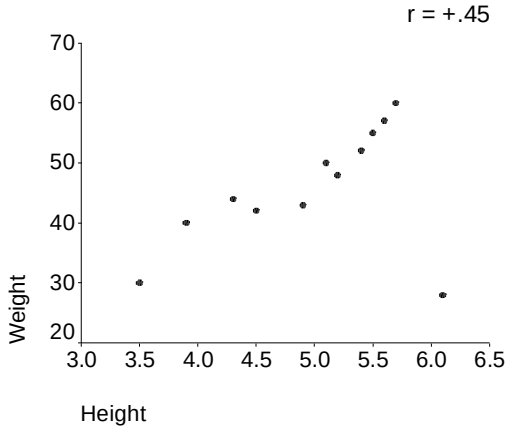
1.4 UNRELIABILITY OF MEASUREMENT

Psychological research involves the use of scales and tests. One of the psychometric property psychological instruments should poses is reliability. Reliability refers to consistency of a measurement. If the instrument is consistent, then the test has high reliability. But at times one of the variable or both the variables may have lower reliability. In this case, the correlation between two less reliable variable reduces. Generally, while interpreting the correlation, the reliability is assumed to be high. The general interpretations of correlations are not valid if the reliability is low. This reduction in the correlation can be adjusted for the reliability of the psychological test. More advanced procedures are available in the books of psychological testing and statistics. They involve calculating disattenuated correlations. Correlation between two variables that have less than perfect reliability is adjusted for unreliability. This is called as disattenuated correlation. If both variables were perfectly reliable then correlation between them is disattenuated correlation.

1.4.1 Outliers

Outliers are extreme score on one of the variables or both the variables. The presence of outliers has deterring impact on the correlation value. The strength and degree of the correlation are affected by the presence of outlier. Suppose you want to compute correlation between height and weight. They are known to correlate positively. Look at the figure below. One of the scores has low score on weight and high score on height (probably, some anorexia patient).

Figure 1.10. Impact of an outlier observation on correlation. Without the outlier, the correlation is 0.95. The presence of an outlier has drastically reduced a correlation coefficient to 0.45.



1.4.2 Curvilinearity

We have already discussed the issue of linearity of the relationship. The Pearson's product moment correlation is appropriate if the relationship between two variables is linear. The relationships are curvilinear then other techniques need to be used. If the degree of cruviliniarity is not very high, high score on both the variable go together, low scores go together, but the pattern is not linear then the useful option is Spearman's *rho*.

1.5 USING RAW SCORE METHOD FOR CALCULATING *r*

The method which we have learned to compute the correlation coefficient is called as deviation scores formula. Now we shall learn another method to calculate Pearson's correlation coefficient. It is called as raw score method. First we will understand how the two formulas are similar. Then we will solve a numerical example for the raw score method. We have learned following formula for calculating *r*.

1.5.1 Formulas for Raw Score

We have already learnt following formula of correlation (eq. 1.8). This is a deviation score formula.

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{nS_X S_Y}$$

The denominator of correlation formula can be written as

$$\sqrt{\sum (X - \bar{X})^2 (Y - \bar{Y})^2} \qquad \text{(eq. 1.10)}$$

Which is

$$\sqrt{(SS_X SS_Y)} \qquad \text{(eq. 1.11)}$$

We have already learnt that

$$SS_X = \sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n} \qquad \text{(eq. 1.12)}$$

and

$$SS_Y = \sum (Y - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} \qquad \text{(eq. 1.13)}$$

The numerator of the correlation formula can be written as

$$\sum (X - \bar{X})(Y - \bar{Y}) = \sum XY - \frac{(\sum X)(\sum Y)}{n} \qquad \text{(eq. 1.14)}$$

So *r* can be calculated by following formula which is a raw score formula:

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{(SS_X SS_Y)}} \qquad \text{(eq. 1.15)}$$

1.5.2 Solved Numerical for Raw Score Formula

We shall solve the same numerical example by using the formulas shown above. Table 3 shows how to calculate Pearson's *r* by using raw score formula.

Table 1.3: Table showing the calculation of r by using raw score formula.

Subject	BHS (X)	BDI (Y)	X^2	Y^2	XY
1	11	13	100	676	260
2	13	16	64	529	184
3	16	14	81	529	207
4	9	10	169	676	338
5	6	8	121	576	264
6	17	16	196	900	420
7	7	9	256	729	432
8	12	12	144	729	324
9	5	7	225	841	435
10	14	15	144	625	300
Summation	110	120	1366	1540	1437
	$\bar{X} = 11$	$\bar{Y} = 12$			

$$SS_X = \sum X^2 - (\sum X)^2 / n = 1366 - (110)^2/10 = 156$$

$$SS_Y = \sum Y^2 - (\sum Y)^2 / n = 1540 - (120)^2 / 10 = 100$$

$$\sum (X - \bar{X})(Y - \bar{Y}) = \sum XY - \frac{(\sum X)(\sum Y)}{n} \qquad = 117$$

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{(SS_X SS_Y)}} \qquad = 0.937$$

Readers might find one of the methods easier. There is nothing special about the methods. One should be able to correctly compute the value of correlation.

1.6 LET US SUM UP

In this unit we started with definition and meaning of correlation and followed it up with how the correlation could be depicted in graphical form and scatter diagram. We then learnt about linear and non-linear and curvilinear relationship amongst variables. We also learnt that the direction of relationship could be either in the positive or in the negative direction. It can also be no correlation in that it can be a zero correlation. Then we learnt the methods of measurement of correlation and

learnt how to use the formula for calculating the Pearson's r, that Pearson's Product Moment Coefficient of Correlation. We then discussed about the building blocks of correlation which included variance and co variance. We also learnt how to test the level of significance of a particular coefficient of correlation calculated by us. Then we learnt about the interpretation of the correlation coefficient and also its ramifications. Then we looked into the unreliability of a correlation and the causes for the same such as the inclusion of outliers etc.

1.7 UNIT END QUESTIONS

1) Problem:

Plot scatter diagram for the following data. Compute Pearson's correlation between x and y. Write the null hypothesis stating that population correlation is zero. Test the significance of the correlation coefficient.

X	Y
12	20
13	22
15	28
17	31
11	22
9	24
8	18
10	21
11	23
7	16

2) Plot scatter for following example. The data was collected on Perceived stress and anxiety on 10 subjects. Compute the Pearson's correlation between them State the null hypothesis. Test the null hypothesis using this hypothesis. Do the similar exercise after deleting a pair that clearly looks an outlier observation.

Perceived stress	Anxiety
9	12
8	11
7	9
4	5
8	9
4	6
6	8
14	2
7	11
11	9
9	11

3) Data showing scores on time taken to complete 200 meters race and duration of practice for 5 runners. Plot the scatter. Compute mean, variance, SD, and covariance. Compute correlation coefficient. Write the null hypothesis.

Time taken (in Seconds)	Duration of Practice (in months)
31	11
32	14
36	9
26	15
38	7

4) Data showing scores on dissatisfaction with work and scores on irritability measured by standardised test for thirteen individuals. Plot the scatter. Compute mean, variance, SD, and covariance. Compute correlation coefficient. Write the null hypothesis stating no relationship. Test the significance at 0.05 level of significance.

Dissatisfaction with work	Irritability scores
12	5
16	7
19	9
27	13
30	16
25	11
22	6
26	14
11	7
17	9
19	14
21	18
23	19

5) Check whether the following statements are true or false.

1)	Positive correlation means as X increases Y decreases.	True/False
2)	Negative correlation means as X decreases Y decreases.	True/False
3)	Generally, in a scatter, lower scores on X are paired with lower scores on Y for negative correlation.	True/False
4)	$-1.00 \leq$ Pearson's correlation $\leq +1.00$	True/False
5)	Generally, in a scatter, lower scores on X are paired with higher scores on Y in positive correlation.	True/False
6)	The scatter diagram cannot indicate the direction of the relationship.	True/False
7)	Percentage of shared variance by X and Y can be obtained by squaring the value of correlation.	True/False

Answers: 1) = False, 2) = False, 3) = False, 4) = True, 5) = False, 6) = False, 7) = True

Answer in brief.

6) What is correlation coefficient?

7) What is the range of correlation coefficient?

8) Is correlation coefficient a percentage?

9) How to calculate common variance from correlation coefficient?

10) What is the percentage of variance shared by X and Y if the $r_{xy} = 0.77$?

11) What is the percentage of variance shared by X and Y if the $r_{xy} = -0.56$?

Answers:

A number expressing the relationship between two variables.

The range of correlation coefficient is from –1.00 to + 1.00.

No. Correlation is not a percentage. But it can be converted into percentage of variance shared.

Common variance is calculated from correlation coefficient by using a formula: $r_{xy}^2 \times 100$.

59.29%

31.36%

1.8 SUGGESTED READINGS

Aron, A., Aron, E. N., Coups, E.J. (2007). *Statistics for Psychology*. Delhi: Pearson Education.

Minium, E. W., King, B. M., & Bear, G. (2001). *Statistical Reasoning in Psychology and Education*. Singapore: John-Wiley.

Guilford, J. P., & Fructore, B. (1978). *Fundamental Statistics for Psychology and Education*. N.Y.: McGraw-Hill.

Wilcoxon, R. R. (1996). *Statistics for Social Sciences*. San Diego: Academic Press.

UNIT 2 OTHER TYPES OF CORRELATION (PHI-COEFFICIENT)

Structure

2.0 Introduction

2.1 Objectives

2.2 Special types of Correlation

2.3 Point Biserial Correlation r_{PB}

- 2.3.1 Calculation of r_{PB}
- 2.3.2 Significance Testing of r_{PB}

2.4 Phi Coefficient (ϕ)

- 2.4.1 Significance Testing of phi (ϕ)

2.5 Biserial Correlation

2.6 Tetrachoric Correlation

2.7 Rank Order Correlations

- 2.7.1 Rank-order Data
- 2.7.2 Assumptions Underlying Pearson's Correlation not Satisfied

2.8 Spearman's Rank Order Correlation or Spearman's rho (r_s)

- 2.8.1 Null and Alternate Hypothesis
- 2.8.2 Numerical Example: for Untied and Tied Ranks
- 2.8.3 Spearman's Rho with Tied Ranks
- 2.8.4 Steps for r_S with Tied Ranks
- 2.8.5 Significance Testing of Spearman's rho

2.9 Kendall's Tau (ô)

- 2.9.1 Null and Alternative Hypothesis
- 2.9.2 Logic of ô and Computation
- 2.9.3 Computational Alternative for ô
- 2.9.4 Significance Testing for ô

2.10 Let Us Sum Up

2.11 Unit End Questions

2.12 Suggested Readings

2.0 INTRODUCTION

We have learned about the correlation as a concept and also learned about the Pearson's coefficient of correlation. We understand that Pearson's correlation is based on certain assumptions, and if those assumptions are not followed or the data is not appropriate for the Pearson's correlation, then what has to be done ? This unit is answering this practical problem. When either the data type or the assumptions are not followed then the correlation techniques listed in this unit are useful. Out of them some are actually Pearson's correlations with different name and some are non-Pearson correlations. The rank data also poses some issues and hence this unit is also providing the answers to this problem. In this unit we shall learn about Special

Types of Pearson Correlation, Special Correlation of Non-Pearson Type, and correlations for rank-order data. The special types of Pearson correlation are Point-Biserial Correlation and Phi coefficient. The non-Pearson correlations are Biserial and Tetrachoric. The rank order correlations discussed are Spearman's *rho* and Kendall's *tau*.

2.1 OBJECTIVES

After completing this unit, you will be able to:

- describe and explain concept of special correlation;
- explain the concept of special correlation and describe and differentiate between their types;
- describe and explain concept of Point-Biserial and Phi coefficient;
- describe and explain concept of Biserial and Tetrachorich coefficient;
- compute and interpret Special correlations;
- test the significance and apply the correlation to the real data;
- explain concept of Spearman's rho and tau coefficient;
- compute and interpret rho and tau; and
- apply the correlation techniques to the real data.

2.2 SPECIAL TYPES OF CORRELATION

The correlation we have learned in the last unit is Pearson's product moment coefficient of correlation. The Pearson's *r* is one of the computational processes for calculating the correlation between two variables. Nevertheless, this is not the only way to calculate correlations. It is just one of the ways of calculating correlation coefficient.

This correlation can be calculated under various restrictions. The variables X and Y were assumed to be continuous variables. The distribution of these variables is expected to be normal. Some homogeneity among the variables is also expected. Linearity of the relationship is also required for computing the Pearson's *r*. There might be instances when one or more of these conditions are not met. In such cases, one needs to use alternative methods of correlations. Some of them are Pearson's correlation modified for specific kind of data. Others are non-Pearson correlations.

Let us take a quick note on distinction between measures of correlation and measures of association. Howell (2002) made this point quite clear. Measures of correlations are those where some sort of order can be assigned to each of the variable. Increment in scores either represent higher levels (or lower levels) of some quantified attribute. For example, number of friends, BHS hope scores, time taken to complete a task, etc. Measures of association are those statistical procedures that are utilised for variables that do not have a property of order. They are categorical variables, or nominal variables, for example association of gender (male and female) with ownership of residence (own and do not own). Both these variables are nominal variables and do not involve any order.

In this unit we shall learn about Special Types of Pearson Correlation, Special Correlation of Non-Pearson Type, and correlations for rank-order data. The special

types of Pearson correlation are Point-Biserial Correlation and Phi coefficient. The non-Pearson correlations are Biserial and Tetrachoric. The rank order correlations discussed are Spearman's *rho* and Kendall's *tau*.

2.3 POINT BISERIAL CORRELATION (r_{PB})

Some variables are dichotomous. The dichotomous variable is the one that can be divided into two sharply distinguished or mutually exclusive categories. Some examples are, male-female, rural-urban, Indian-American, diagnosed with illness and not diagnosed with illness, Experimental group and Control Group, etc. These are the truly dichotomous variables for which no underlying continuous distribution can be assumed. Now if we want to correlate these variables, then applying Pearson's formula have problems because of lack of continuity. Pearson's correlation requires continuous variables.

Suppose we are correlating gender, then male will be given a score of 0, and females will be given a score of 1 (or vice versa; indeed you can give a score of 5 to male and score of 11 to female and it won't make any difference for the correlation calculated).

Point Biserial Correlation (r_{pb}) is Pearson's Product moment correlation between one truly dichotomous variable and other continuous variable. Algebraically, the $r_{pb} = r$. So we can calculate r_{pb} in a similar way.

2.3.1 Calculation of r_{pb}

Let's look at the following data. It is a data of 20 subjects, out of which 9 are male and 11 are females. Their marks in the final examination are also provided. We want to correlate marks in the final examination with sex of the subject. The marks obtained in the final examination are a continuous variable whereas sex is truly dichotomous variable, taking two values male or female. We are using value of 0 for male subject and value of 1 for female subjects. The correlation appropriate for this purpose is Point-Biserial correlation (r_{pb}).

Table 1: Data showing the gender and mark for 20 subjects

Subject	Sex (male) X	Marks (Y)	Subject	Sex (Female (X)	Marks (Y)
1	0	46	11	1	58
2	0	74	12	1	69
3	0	58	13	1	76
4	0	67	14	1	78
5	0	62	15	1	65
6	0	71	16	1	69
7	0	54	17	1	59
8	0	63	18	1	53
9	0	53	19	1	73
10	1	67	20	1	81

$\text{Mean}_{sex} = 0.55$ $\text{Mean}_{marks} = 64.8$ $\text{Mean Marks}_{male} = 60.88$

$S_{sex} = 0.497$ $S_{marks} = 9.17$ $\text{Mean Marks}_{female} = 68$

$Cov_{XY} = 1.76$

$$r = \frac{Cov_{XY}}{S_X S_Y} = \frac{1.76}{0.497 \times 9.17} = 0.386$$

The Pearson's correlation (point biserial correlation) between sex and marks obtained is 0.386. The sign is positive. The sign is arbitrary and need to be interpreted depending on the coding of the dichotomous group. The interpretation of the sign is the group that is coded as 1 has a higher mean than the group that is coded as 0. The strength of correlation coefficient is calculated in a similar way. The correlation is 0.386, so the percentage of variance shared by both the variables is r^2 for Pearson's correlation. Same would hold true for point biserial correlation. The r_{pb}^2 is $0.386^2 = 0.149$. This means that 15% of information in marks is shared by sex.

2.3.2 Significance Testing of r_{pb}

The null hypothesis and alternative hypothesis for this purpose are as follows:

H_o: $\tilde{n} = 0$

H_A: $\tilde{n} \neq 0$

Since the r_{pb} is Pearson's correlation, the significance testing is also similar to it. the *t*- distribution is used for this purpose with $n - 2$ as *df*.

$$t = \frac{r_{pb}\sqrt{n-2}}{\sqrt{1-r_{pb}^2}} \qquad \text{(eq. 2.1)}$$

The t value for our data is 1.775. The $df = n - 2 = 20 - 2 = 18$. The value is not significant at 0.05 level. Hence we retain the null hypothesis.

2.4 PHI COEFFICIENT (ϕ)

The Pearson's correlation between one dichotomous variable and another continuous variable is called as point-biserial correlation. When both the variables are dichotomous, then the Pearson's correlation calculated is called as Phi Coefficient (ϕ).

For example, let us say that you have to compute correlation between gender and ownership of the property. The gender takes two levels, male and female. The ownership of property can be measured as either the person owns a property and the person do not own property. Now you have both the variables measured as dichotomous variables. Now if you compute the Pearson's correlation between these two variables is called as Phi Coefficient (ϕ). Both the variables take value of either of 0 or one. Look at the data given in the table below.

Table 2: Data and calculation for correlation between gender and ownership of property

X: Gender	0= Male 1 = Female											
Y: Ownership of Property	0=No ownership 1 = Ownership											
X	1	0	1	1	0	0	0	0	1	1	1	0
Y	0	1	0	1	1	1	0	1	0	0	1	1

Calculations

$\overline{X} = 0.5$ $S_x = 0.52$

$\overline{Y} = .58$ $S_y = 0.51$ Cov $_{XY}$ = – 0.13

$$r_{XY} = \phi_{XY} = \frac{Cov_{XY}}{S_X S_Y} = \frac{-0.13}{0.52 \times 0.51} = -.465$$

The value of ϕ coefficient is found to be – .465.

The relationship is negative, is function of the way we have assigned the number 0 and 1 to each of the variable. If we assign 0 to females and 1 to males, then we will get the same value of correlation with positive sign. Nevertheless, this does not mean that sign of the relationship cannot be interpreted. Once these numbers have been assigned, then we can interpret the sign. Male is 0 and female is 1; whereas 0 = no ownership and 1 is ownership.

The negative relation can be interpreted as follows: as we move from male to female we move negatively from no ownership to ownership. Meaning that male have more ownership than females. We can also calculate the proportion of variance shared by these two variables.

That is $r^2 = \phi^2 = -.465^2 = 0.216$ percent.

2.4.1 Significance Testing of Phi (ϕ)

The significance can be tested by using the Chi-Square ($\div^2$) distribution.

The ϕ can be converted into the $\div^2$ by obtaining a product of n and ϕ^2.

The Chi-Square of $n\phi^2$ will have $df = 1$.

The null and alternative hypothesis are as follows:

H_O: $\tilde{n} = 0$

H_A: $\tilde{n}$ ‘‘‘ 0

$\div^2 = n\phi^2 = 12 \times .216 = 2.59$ (eq. 2.2)

The value of the chi-square at 1 *df* is 3.84. the obtained value is less than the tabled value. So we accept the null hypothesis which states that the population correlation is zero.

One need to know that this is primarily because of the small sample size. If we take a larger sample, then the values would be significant. Quickly note the relationship between $\div^2$ and ϕ.

$$\phi = \sqrt{\frac{\chi^2}{n}} \qquad \text{(eq. 2.3)}$$

So one can compute the chi-square and then calculate the phi coefficient.

2.5 BISERIAL CORRELATION

The biserial correlation coefficient (r_b), is a measure of correlation. It is like the point-biserial correlation. But point-biserial correlation is computed while one of the variables is dichotomous and do not have any underlying continuity. If a variable has underlying continuity but measured dichotomously, then the biserial correlation can be calculated.

An example might be mood (happy-sad) and hope, which we have discussed in the first unit. Suppose we measure hope with BHS and measure mood by classifying those who have clinically low *vs*. normal mood. Actually, it is fair to assume that mood is a normally distributed variable.

But this variable is measured discretely and takes only two values, low mood (0) and normal mood (1).

Let's call continuous variable as Y and dichotomized variable as X. the values taken by X are 0 and 1.

So biserial correlation is a correlation coefficient between two continuous variables (X and Y), out of which one is measured dichotomously (X). The formula is very similar to the point-biserial but yet different:

$$r_b = \left[\frac{\overline{Y}_1 - \overline{Y}_0}{S_Y}\right]\left[\frac{P_0 P_1}{h}\right] \qquad \text{(eq. 2.4)}$$

where $\overline{Y}_0$ and $\overline{Y}_1$ are the Y score means for data pairs with an X score of 0 and 1, respectively, P_0 and P_1 are the proportions of data pairs with X scores of 0 and 1, respectively, and S_Y is the standard deviation for the Y data, and *h* is ordinate or the height of the standard normal distribution at the point which divides the proportions of P_0 and P_1.

The relationship between the point-biserial and the biserial correlation is as follows.

$$r_b = \frac{r_{pb}\sqrt{p_0 p_1}}{h} \qquad \text{(e.q 2.5)}$$

So once you compute the r_{pb}, its easy to compute the r_b.

2.6 TETRACHORIC CORRELATION (r_{TET})

Tetrachoric correlation is a correlation between two dichotomous variables that have underlying continuous distribution. If the two variables are measured in a more refined way, then the continuous distribution will result. For example, attitude to females and attitude towards liberalisation are two variables to be correlated. Now, we simply measure them as having positive or negative attitude. So we have 0 (negative attitude) and 1 (positive attitude) scores available on both the variables. Then the correlation between these two variables can be computed using Tetrachoric correlation (r_{tet}).

The correlation can be expressed as

$$r = \cos\theta \qquad \text{(eq. 2.6)}$$

Where, è is angle between the vector X and Y. Using this logic, r_{tet} can also be calculated.

$$r_{tet} = \cos\left[\frac{180^0}{1+\sqrt{\frac{ad}{bc}}}\right] \qquad \text{(eq. 2.6)}$$

Look at the following data summarised in table. 3.

Table 3: Data for Tetrachoric correlation.

		X variable: Attitude towards women		
		0 (Negative attitude)	1 (Positive attitude)	Sum of row
Attitude towards Liberalisation	0 (Negative attitude)	68 (a)	32 (b)	100
	1 (Positive attitude)	30 (c)	70 (d)	100
	Sum of columns	98	102	total =200

The table values are self explanatory. Out of 200 individuals, 68 have negative attitude towards both variables, 32 have negative attitude to liberalisation but positive attitude to women, and so on. The tetrachoric correlation can be computed as follows.

$$r_{tet} = \cos\left[\frac{180^0}{1+\sqrt{\frac{ad}{bc}}}\right] = \cos\left[\frac{180^0}{1+\sqrt{\frac{(68)(70)}{(30)(32)}}}\right] = \cos 55.784^0 = .722$$

So the tetrachoric correlation between attitude towards liberalisation and attitude towards women is positive.

2.7 RANK ORDER CORRELATIONS

We have learned about Pearson's correlation in earlier unit. The Pearson's correlation is calculated on continuous variables. Pearson's correlation is not advised under two circumstances: one, when the data are in the form of ranks and two, when the assumptions of Pearson's correlation are not followed by the data. In this condition, the application of Pearson's correlations is doubtful. Under such circumstances, rank-order correlations constitute one of the important options. The ordinal scale data is called as rank-order data. Now let us look at these two aspects, rank-order and assumption of Pearson's correlations, in greater detail.

2.7.1 Rank-Order Data

When the data is in rank-order format, then the correlation that can be computed is called as rank order correlations. The rank-order data present the ranks of the individuals or subjects. The observations are already in the rank order or the rank order is assigned to them. Marks obtained in the unit test will constitute a continuous data. But if only the merit list of the students is displayed then the data is called as rank order data. If the data is in terms of ranks, then Pearson's correlation need not be done. Spearman's rho constitutes a good option.

2.7.2 Assumptions Underlying Pearson's Correlation not Satisfied

The statistical significance testing of the Pearson's correlation requires some assumptions about the distributional properties of the variables. We have already delineated these assumptions in the earlier unit. When the assumptions are not followed by the data, then employing the Pearson's correlation is problematic. It should be noted that small violations of the assumptions does not influence the distributional properties and associated probability judgments. Hence it is called as a robust statistics. However, when the assumptions are seriously violated, then application of Pearson's correlation should no longer be considered as a choice. Under such circumstances, Rank order correlations should be preferred over Pearson's correlation.

It needs to be noted that rank-order correlations are applicable under the circumstances when the relationship between two variables is not linear but still it is a *monotonic* relationship. The *monotonic* relationship is one where values in the data are consistently increasing and never decreasing or consistently decreasing and never increasing. Hence, monotonic relationship implies that as X increases Y consistently increase or as X increases Y consistently decrease. In such cases, rank-order is a better option than Pearson's correlation coefficient.

However, some caution should be observed while doing so. A careful scrutiny of Figure 1 below indicates that, in reality, it is a power function. So actually a relationship between X and Y is not linear but curvilinear power function. So, indeed, curve-fitting is a best approach for such data than using the rank order correlation.

The rank-order can be used with this data since the curvilinear relationship shown in figure 1 is also a *monotonic* relationship. It must be kept in mind that all curvilinear relationships would not be monotonic relationships.

In the previous unit, we have discussed the issue of linearity in the section 1.1.2. I have exemplified the non-linear relationship with Yorkes- Dodson law. It states that relationship between stress and performance is non-linear relationship. But this relationship is NOT a monotonic relationship because initially Y increasers with the corresponding increase in X. But beyond the modal value of X, the scores on Y decrease. So this is not a monotonic relationship. Hence, rank-order correlations should not be Calculated for such a data.

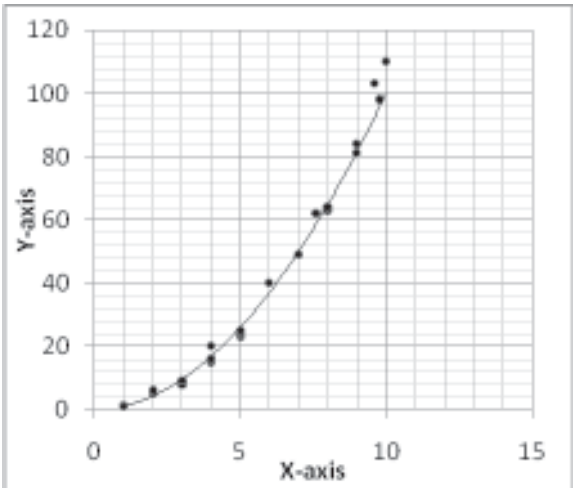


Fig. 1: The figure shows a monotonic relationship between X and Y

2.8 SPEARMAN'S RANK-ORDER CORRELATION OR SPEARMAN'S *RHO* (r_S)

A well-known psychologist and intelligence theorist, Charles Spearman (1904), developed a correlation procedure called in his honor as Spearman's rank-order correlation or Spearman's *rho* (r_s). It was developed to compute correlation when the data is presented on two variables for *n* subjects. It can also be calculated for data of *n* subjects evaluated by two judges for inter-judge agreement. It is suitable for the rank-order data. If the data on X or Y or on both the variables are in rank-order then Spearman's *rho* is applicable. It can also be used with continuous data when the assumptions of Pearson's assumptions are not satisfied. It is used to assess a monotonic relationship.

The range of Spearman's *rho* (r_s) is also from – 1.00 to + 1.00. Like Pearson's correlation, the interpretation of Spearman's *rho* is based on sign of the coefficient and the value of the coefficient.

If the sign of r_s is positive the relationship is positive, if the sign of r_s is negative then the relationship is negative. If the value of r_s is close to zero then relationship is weak, and as the value of r_s approaches to ± 1.00, the strength of relationship increases. When the value of r_s is zero then there is no relationship between X and Y. If r_s is ± 1.00, then the relationship between X and Y is perfect. Whatever the value of r_s may take, it does not directly imply causation. We have already discussed the correlation and causality in the previous unit.

2.8.1 Null and Alternative Hypothesis

The Spearman's *rho* can be computed as a descriptive statistics. We do not carry out statistical hypothesis testing for descriptive use of *rho*. If the r_s is computed as a statistic to estimate population correlation (parameter), then null and alternative hypothesis are required.

The null hypothesis states that

H_O: $\tilde{n}_s = 0$

It means that the value of Spearman's correlation coefficient between X and Y is zero in the population represented by sample.

The alternative hypothesis states that

H_A: $\tilde{n}_s \neq 0$

It means that the value of Spearman's *rho* between X and Y is not zero in the population represented by sample. This alternative hypothesis would require a two-tailed test.

Depending on the theory, the other alternatives could also be written. They are either

H_A: $\tilde{n}_s < 0$

or

H_A: $\tilde{n}_s > 0$.

The first alternative hypothesis, H_A, states that the population value of Spearman's *rho* is smaller than zero. The second H_A denotes that the population value of Spearman's *rho* is greater than zero. Remember, only one of them has to be tested and not both.

You can recall from earlier discussion that one-tailed test is required for this hypothesis.

2.8.2 Numerical Example: for Untied and Tied Ranks

Very obviously, the data on X and Y variables on are required to compute Spearman's *rho*. If the data are on continuous variables then it need to be converted into a rank-orders. The computational formula of Spearman's *rho* (r_s) is as follows:

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} \qquad \text{(eq. 2.7)}$$

Where,

r_s = Spearman's rank-order correlation

D = difference between the pair of ranks of X and Y

n = the number of pairs of ranks

Steps:

Let's solve an example. We have to appear for entrance examination after the under-graduate studies. We are interested in correlating the undergraduate marks and performance in the entrance test. We have a data of 10 individuals. But we only have ranks of these individuals in undergraduate examination, and merit list of the entrance performance. We want to find the correlation between rank in undergraduate examination and rank in entrance. The data are provided in table 4 and 5. Since this is a rank order data, we can carry out the Spearman's *rho*. (If the data on one or both variable were continuous, we need to transfer this data into ranks for computing the Spearman's *rho*.)

Table 4: Data for Spearman's rho.

Students	Rank in Undergraduate Examination (X)	Rank in entrance test (Y)
A	1	4
B	5	6
C	3	2
D	6	7
E	9	10
F	2	1
G	4	3
H	10	9
I	8	8
J	7	5

The steps for computation of r_s are given below:

Step 1: List the names/serial number of subjects (students, in this case) in column 1.

Step 2: Write the scores of each subject on X variable (undergraduate examination) in the column labeled as X (column 2), and write the scores of each subject on Y variable (Entrance test) in the column labeled as Y (column 3). We will skip this step because we do not have original scores in undergraduate examination and entrance test.

Step 3: Rank the scores of X variable in ascending order. Give rank 1 to the lowest score, 2 to the next lowest score, and so on. In case of our data, the scores are already ranked.

Step 4: Rank the scores of Y variable in ascending order. Give rank 1 to the lowest score, 2 to the next lowest score, and so on. This column is labeled as R_Y (Column 5). Do cross-check your ranking by calculating the sum of ranks. In case of our data, the scores are already ranked.

Step 5: Now find out D, where $D = R_X - R_Y$ (Column 6).

Step 6: Square each value of D and enter it in the next column labeled as D^2 (Column7). Obtain the sum of the D^2. This is written at the end of the column D^2. This $\sum D^2$ is 18 for this example.

Step 7: Use the equation 4.1 (given below) to compute the correlation between rank in undergraduate examination and rank in entrance test.

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} \qquad \text{(eq. 2.8)}$$

Table 5: Table showing the data on rank obtained in undergraduate examination and ranks in entrance examination. It also shows the computation of Spearman's *rho*.

Students	Rank in Undergraduate Examination (X)	Rank in entrance test (Y)	R_X	R_Y	$D = R_X - R_Y$	D^2
A	1	4	1	4	-3	9
B	5	6	5	6	-1	1
C	3	2	3	2	1	1
D	6	7	6	7	-1	1
E	9	10	9	10	-1	1
F	2	1	2	1	1	1
G	4	3	4	3	1	1
H	10	9	10	9	1	1
I	8	8	8	8	0	0
J	7	5	7	5	2	4
$n = 10$						$\sum D^2=20$

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6\times 20}{10(10^2-1)} = 1 - \frac{180}{990} = 1 - 0.1818 = 0.818$$

Now the Spearman's *rho* has been computed for this example. The value of *rho* is 0.818. This value is positive value. It shows that the correlation between the ranks in undergraduate examination and the ranks in entrance test is positive. It indicates that the relationship between them is positively monotonic. The value of the correlation coefficient is very close to 1.00 which indicates that the strength association between the two set of ranks is very high. The tied ranks were not employed in this example since it was the first example. Now I shall introduce you to the problem of tied ranks.

Interesting point need to be noted about the relationship between Pearson's correlation

and Spearman's *rho*. The Pearson's correlation on ranks of X and Y (i.e., R_X and R_Y) is equal the Spearman's *rho* on X and Y. That's the relationship between Pearson's *r* and Spearman's *rho*. The Spearman's *rho* can be considered as a special case of Pearson's *r*.

2.8.3 Spearman's *rho* with Tied Ranks

The ranks are called as *tied ranks* when two or more subjects have the same score on a variable. We usually get larger than the actual value of Spearman's *rho* if we employ the formula in the equation 2.1 for the data with the tied ranks. So the formula in equation 2.1 is not appropriate for tied ranks. A correction is required in this formula in order to calculate correct value of Spearman's *rho*. The recommended procedure of correction for tied ranks is computationally tedious. So we shall use a computationally more efficient procedure. The easier procedure of correction actually uses Pearson's formula on the ranks. The formula and the steps are as follows:

$$r = r_s = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left[\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right]\left[\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right]}} \qquad \text{(eq. 2.10)}$$

Where,

r_s = Spearman's *rho*

X = ranks of variable X

Y = rank on variable Y

n = number of pairs

Look at the example we have solved for Pearson's correlation. It is an example of relationship between BHS and BDI. The data is different than the one we have used in the earlier unit. We shall solve this example with Spearman's *rho*.

2.8.4 Steps for r_s with Tied Ranks

If the data are not in ranks, then convert it into rank-order. In this example, we have assigned ranks to X and Y (column 2 and 3) in column 4, and 5.

Appropriately rank the ties (Cross-check the ranking by using sum of ranks check). This is the basic information for the Spearman's *rho*.

Compute the square of rank of X and rank of Y for all the observations. It is in columns 6 and 7.

Multiply the rank of X by rank of Y for each observation. It is provided in column 8.

Obtain sum of all the columns. Now all the basic data for the computation is available.

Enter this data into the formula shown in the equation 2.2 and calculate r_s.

Table 6: Spearman's *rho* for tied ranks

Subject	BHS (X)	BDI (Y)	Rank X	Rank Y	(Rank X)^2	(Rank Y)^2	(Rank X) (Rank Y)
1	7	8	3.5	2.5	12.25	6.25	8.75
2	11	16	6.5	9.5	42.25	90.25	61.75
3	16	14	9	7	81	49	63
4	9	12	5	5.5	25	30.25	27.5
5	6	8	2	2.5	4	6.25	5
6	17	16	10	9.5	100	90.25	95
7	7	9	3.5	4	12.25	16	14
8	11	12	6.5	5.5	42.25	30.25	35.75
9	5	7	1	1	1	1	1
10	14	15	8	8	64	64	64
Sum			55	55	384	383.5	375.75

$$r_s = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left[\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right]\left[\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right]}} = \frac{375.75 - \frac{(55)(55)}{10}}{\sqrt{\left[384 - \frac{55^2}{10}\right]\left[383.5 - \frac{55^2}{10}\right]}} = \frac{73.25}{81.2496} = 0.902$$

The Spearman's *rho* for this example is 0.902. Since this is a positive value, the relationship between them is also positive. This value is rather near to 1.00. So the strength of association between the ranks of BDI and BHS are very high. This is a simpler way to calculate the Spearman's *rho* with tied ranks. Now, we shall look at the issue of significance testing of the Spearman's *rho*.

2.8.5 Significance Testing of Spearman's *rho*

Once the statistics of Spearman's *rho* is calculated, then the significance of Spearman's *rho* need to be found out. The null hypothesis tested is

H_O: $\tilde{n}_s = 0$

It states that the value Spearman's *rho* between X and Y is zero in the population represented by sample.

The alternative hypothesis is

H_A: $\tilde{n}_s \neq 0$

It states that the value Spearman's *rho* between X and Y is not zero in the population represented by sample. This alternative hypothesis requires a two-tailed test. We have already discussed about writing a directional alternative which requires one-tailed test.

We need to refer to Appendix D for significance testing. The appendix in statistics book, provides critical values for one-tailed as well as two-tailed tests. Let us use the table for the purpose of hypothesis testing for the first example of correlation between ranks in undergraduate examination and entrance test (table 2).

The obtained Spearman's *rho* is 0.818 on the sample of 10 individuals. For $n = 10$, and two-tailed level of significance of 0.05, the critical value of $r_s = 0.648$. The critical value of $r_s = 0.794$ at the two-tailed significance level of 0.01.

The obtained value of 0.818 is larger than the critical value at 0.01. So the obtained correlation is significant at 0.01 level (two-tailed). We reject the null hypothesis and accept the alternative hypothesis. It indicates that the value of the Spearman's *rho* is not zero in the population represented by the sample.

For the second example (table 3), the obtained r_s value is 0.902 on the sample of 10 individuals. For $n = 10$, the critical value is 0.794 at the two-tailed significance level of 0.01. The obtained value of 0.902 is larger than the critical value at 0.01. So the obtained correlation is significant at 0.01 level (two-tailed). Hence, we reject the null hypothesis and accept the alternative hypothesis.

When the sample size is greater than ten, then the *t*-distribution can be used for computing the significance with $df = n - 2$. Following equation is used for this purpose.

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \qquad \text{(eq. 2.10)}$$

For the example shown in table 2, the *t*-value is computed using equation 2.11.

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} = \frac{0.818\sqrt{10-2}}{\sqrt{1-0.818^2}} = 4.022 \qquad \text{(eq.2.11)}$$

At the $df = 10 - 2 = 8$, the critical *t*-value at 0.01 (two-tailed) is 3.355. The obtained *t*-value is larger than the critical *t*-value. Hence, we reject the null hypothesis and accept the alternative hypothesis.

2.9 KENDALL'S *TAU* (τ)

Kendall's *tau* is another useful measure of correlation. It is as an alternative to Spearman's *rho* (r_s).

This correlation procedure was developed by Kendall (1938). Kendall's *tau* is based on an analysis of two sets of ranks, X and Y. Kendall's *tau* is symbolised as ô, which is a lowercase Greek letter *tau*. The parameter (population value) is symbolised as ô and the statistics computed on the sample is symbolised as. The range of *tau* is from – 1.00 to + 1.00. The interpretation of *tau* is based on the sign and the value of coefficient. The *tau* value closer to ±1.00 indicates stronger relationship. Positive value of *tau* indicates positive relationship and vice versa. It should be noted that Kendall's Concordance Coefficient is a different statistics and should not be confused with Kendall's *tau*.

2.9.1 Null and Alternative Hypothesis

When the Kendall's *tau* is computed as a descriptive statistics, statistical hypothesis testing is not required. If the sample statistic $\tilde{\tau}$ is computed to estimate population correlation ô, then null and alternative hypothesis are required.

The null hypothesis states that

H_O: ô = 0

It stated that the value Kendall's *tau* between X and Y is zero in the population represented by sample.

The alternative hypothesis states that

H_A: $\hat{o} \neq 0$

It states that the value Kendall's *tau* between X and Y is not zero in the population represented by sample. This alternative hypothesis requires a two-tailed test.

Depending on the theory, the other alternatives could be written. They are either

1) H_A: $\hat{o} < 0$ or

2) H_A: $\hat{o} > 0$.

The first H_A denotes that the population value of Kendall's *tau* is smaller than zero.

The second H_A denotes that the population value of Kendall's *tau* is greater than zero. Remember, only one of them has to be tested and not both. One-tailed test is required for these hypotheses.

2.9.2 Logic of τ and Computation

The *tau* is based on concordance and discordance among two sets of ranks. For example, table 4.4 shows ranks of four subjects on variables X and Y as R_X and R_Y. In order to obtain concordant and discordant pairs, we need to order one of the variables according to the ranks, from lowest to highest (we have ordered X in this fashion).

Take a pair of ranks for two subjects A (1,1) and B (2,3) on X and Y.

Now, if sign or the direction of $R_X - R_X$ for subject A and B is similar to the sign or direction of $R_Y - R_Y$ for subject A and B, then the pair of ranks is said to be concordant (i.e., in agreement).

In case of subject A and B, the $R_X - R_X$ is (1 – 2 = – 1) and $R_Y - R_Y$ is also (1 – 3 = – 2). The sign or direction of A and B pair is in agreement. So pair A and B is called as concordant pair.

Look at second example of B and C pair. The $R_X - R_X$ is (2 – 3 = – 1) and $R_Y - R_Y$ is also (3 – 2 = +1). The sign or the direction of B and C pair is not in agreement. This pair is called as discordant pair.

Table 7: Small data example for *tau* on four subjects

Subject	R_X	R_Y
A	1	1
B	2	3
C	3	2
D	4	4

How many such pair we need to evaluate? They will be $n(n-1)/2 = (4 \times 3)/2 = 6$, so six pairs. Here is an illustration: AB, AC, AD, BC, BD, and CD. Once we know the concordant and discordant pairs, then we can calculate by using following equation.

$$\tilde{\tau} = \frac{n_C - n_D}{\left[\frac{n(n-1)}{2}\right]} \qquad \text{(eq. 2.13)}$$

Where,

$\tilde{\tau}$ = value of ô obtained on sample

n_C = number of concordant pairs

n_D = number of discordant pairs

n = number of subjects

Now, I illustrate a method to obtain the number of concordant (n_C) and discordant (n_D) pairs for this small data in the table above. We shall also learn a computationally easy method later.

Step 1. First, Ranks of X are placed in second row in the ascending order.

Step 2. Accordingly ranks of Y are arranged in the third row.

Step 3. Then the ranks of Y are entered diagonally.

Step 4. Start with the first element in the diagonal which is 1 (row 4).

Step 5. Now move across the row.

Step 6. Compare it (1) with each column element of Y. If it is smaller then enter C in the intersection. If it is larger, then enter D in the intersection. For example, 1 is smaller than 3 (column 3) so C is entered.

Step 7. In the next row (row 5), 3 is in the diagonal which is greater than 2 (column 4) of Y, so D is entered in the intersection.

Step 8. Then "C and "D are computed for each row.

Step 9. The n_C is obtained from $\sum\sum$C (i.e., 5) and

Step 10. n_D is obtained from $\sum\sum$D (i.e., 1).

Step 11. These values are entered in the equation 4.4 to obtain.

Table 8. Computation of concordant and discordant pairs.

Subjects	A	B	C	D	∑C	∑D
Rank of X	1	2	3	4		
Rank of Y	1	3	2	4		
	1	C	C	C	3	0
		3	D	C	1	1
			2	C	1	0
				4	0	0
					∑∑C = 5	∑∑D = 1

$$\tilde{\tau} = \frac{n_C - n_D}{\left[\frac{n(n-1)}{2}\right]} = \frac{5-1}{\left[\frac{4(4-1)}{2}\right]} = \frac{4}{6} = 0.667$$

2.9.3 Computational Alternative for τ

This procedure of computing the *tau* is tedious. I suggest an easier alternative. Suppose, we want to correlate rank in practice sessions and rank in sports competitions. We also know the ranks of the sportspersons on both variables. The data are given below for 10 sportspersons.

Table 9: Data of 10 subjects on X (rank in practice session) and Y (ranks in sports competition)

	Subjects being ranked									
	A	B	C	D	E	F	G	H	I	J
Practice session (Ranks on X)	1	2	3	4	5	6	7	8	9	10
Sports competition (Ranks on Y)	2	1	5	3	4	6	10	8	7	9

First we arrange the ranks of the students in ascending order (in increasing order; begin from 1 for lowest score) according to one variable, X in this case. Then we arrange the ranks of Y as per the ranks of X. I have drawn the lines to connect the comparable ranking of X with Y. Please note that lines are not drawn if the subject gets the same rank on both the variables. Now we calculate number of inversions. Number of inversions is number of intersection of the lines. We have five intersections of the lines.

So the following equation can be used to compute $\tilde{\tau}$

$$\tilde{\tau} = 1 - \frac{2(n_s)}{\frac{n(n-1)}{2}} \qquad \text{(eq. 2.14)}$$

Where

$\tilde{\tau}$ = sample value of ô

n_s = number of inversions

n = number of subjects

$$\tilde{\tau} = 1 - \frac{2(n_s)}{\frac{n(n-1)}{2}} = 1 - \frac{2(5)}{\frac{10(10-1)}{2}} = 1 - \frac{10}{45} = 1 - 0.222 = 0.778$$

The value of Kendall's *tau* for this data is 0.778. The value is positive. So the relationship between X and Y is positive. This means as the rank on time taken increases the rank on subject increases. Interpretation of *tau* is straightforward. For example, if the $\tilde{\tau}$ is 0.778, then it can be interpreted as follows: if the pair of subjects is sampled at random, then the probability that their order on two variables (X and Y) is similar is 0.778 higher than the probability that it would be in reverse order. The calculation of *tau* need to be modified for tied ranks. Those modifications are not discussed here.

2.9.4 Significance Testing of ô

The statistical significance testing of Kendall's *tau* is carried out by using either Appendix E and referring to the critical value provided in the Appendix E. The other way is to use the z transformation. The z can be calculated by using following equation

$$z = \frac{\tilde{\tau}}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \qquad \text{(eq. 2.15)}$$

You will realise that the denominator is the standard error of *tau*. Once the Z is calculated, you can refer to Appendix A for finding out the probability.

For our example in table 4, the value of $\tilde{\tau} = 0.664$ for the $n = 4$. The Appendix E provides the critical value of 1.00 at two-tailed significance level of 0.05. The obtained value is smaller than the critical value. So it is not statistically significant. Hence, we retain the null hypothesis which states H_O: ô = 0. So we accept this hypothesis. It implies that the underlying population represented by the sample has no relationship between X and Y.

For example in table 6, the obtained value of *tau* is 0.778 with the $n = 10$. From the Appendix E, for the $n = 10$, the critical value of *tau* is 0.644 at two-tailed 0.01 level of significance. The value obtained is 0.778 which is higher than the critical value of 0.664. So the obtained value of *tau* is significant at 0.01 level. Hence, we reject the null hypothesis H_O: ô = 0 and accept the alternative hypothesis H_A: ô ≠ 0. It implies that the value of *tau* in the population represented by sample is other than zero. So there exists a positive relationship between practice ranks and sports competition ranks.

Other way of testing significance is to convert the obtained value of the tau into z. Then use the z distribution for testing the significance of the tau. For this purpose, following formula can be used.

$$z = \frac{\tilde{\tau}}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} = \frac{0.778}{\sqrt{\frac{2(2\times10+5)}{9\times10(10-1)}}} = 3.313$$

The z table (normal distribution table) in the Appendix A has a value of z = 1.96 at 0.05 level and 2.58 at 0.01 level. The obtained value of z = 3.313 is far greater than these values. So we reject the null hypothesis at 0.01 level of significance.

Kendall's *tau* is said to better alternative to Spearman's *rho* under the conditions of tie ranks. The *tau* is also supposed to do better than Pearson's *r* under the conditions of extreme non-normality. This holds true only under the conditions of very extreme cases. Otherwise, Pearson's *r* is still a coefficient of choice.

2.10 LET US SUM UP

In this unit, we have learned the specific types of correlations that can be used under circumstances that are special. These correlations are either Pearsons correlations with different names or non-Pearson correlations. We have also learned to compute the values as well as test the significances of these correlations. We have also learned the coorelations that can be calculated for the ordinal data. They are Spearman's rho and tau. Indeed we also got to know that Spearman's rho can be considered as a special case of Pearson's correlation. The tau is useful under ties ranks. This will help you to handle the correlation data of various types.

2.11 UNIT END QUESTIONS

1) What are the special types of correlations and why are they to be used?

2) Discuss the point biserial correlation and indicate its advantages.

3) Calculate point biserial for the following data:

Subject	Sex (male) X	Marks (Y)	Subject	Sex (Female (X)	Marks (Y)
1	0	30	11	1	38
2	0	56	12	1	69
3	0	68	13	1	78
4	0	48	14	1	58
5	0	52	15	1	55
6	0	80	16	1	89
7	0	78	17	1	82
8	0	72	18	1	85
9	0	55	19	1	73
10	0	48	20	1	62

4) How will you do the significance of testing for point biserial correlation

5) When do we use Phi Coefficient?

6) Calculate phi coefficient for the following data

X: Gender — 0= Male, 1 = Female

Y: Ownership of Property — 0=No ownership, 1 = Ownership

X	1	0	1	1	0	1	1	0	0	1	1	0
Y	1	1	0	0	1	0	0	1	1	0	1	1

7) What is biserial correlation? When do we use biserial correlation?

8) Discuss the use of Tetrachoric correlation.

9) What are the important assumptions of rank order correlation?

10) Discuss in detail Spearman's Rank Correlation and compare it with Kendall's tau.

11) Calculate Rho for the following data and test the significance of Rho

Students	Marks in history	Marks in English
A	50	60
B	45	48
C	63	72
D	65	76
E	48	58
F	59	60
G	62	68

12) Discuss Kendall's Tau.

13) Discuss the significance testing of Tau.

14) Calculate Tau for the following data

	A	B	C	D	E	F	G	H	I	J
Practice session (Ranks on X)	1	2	3	4	5	6	7	8	9	10
Sports competition (Ranks on Y)	5	1	2	4	4	10	6	7	9	8

2.12 SUGGESTED READINGS

Garrett, H.E. (19). *Statistics In Psychology And Education*. Goyal Publishing House, New Delhi.

Guilford, J.P.(1956). *Fundamental Statistics in Psychology and Education.* McGraw Hill Book company Inc. New York.

UNIT 3 PARTIAL AND MULTIPLE CORRELATIONS

Structure

3.0 Introduction

3.1 Objectives

3.2 Partial Correlation (r_p)

3.2.1 Formula and Example

3.2.2 Alternative Use of Partial Correlation

3.3 Linear Regression

3.4 Part Correlation (Semipartial correlation) r_{sp}

3.4.1 Semipartial Correlation: Alternative Understanding

3.5 Multiple Correlation Coefficient (R)

3.6 Let Us Sum Up

3.7 Unit End Questions

3.8 Suggested Readings

3.0 INTRODUCTION

While learning about correlation, we understood that it indicates relationship between two variables. Indeed, there are correlation coefficients that involve more than two variables. It sounds unusual and you would wonder how to do it? Under what circumstance it can be done? Let me give you two examples. The first is about the correlation between cholesterol level and bank balance for adults. Let us say that we find a positive correlation between these two factors. That is, as the bank balance increases, cholesterol level also increases. But this is not a correct relationship as Cholesterol level can also increase as age increases. Also as age increases, the bank balance may also increase because a person can save from his salary over the years. Thus there is age factor which influences both cholesterol level and bank balance. Suppose we want to know only the correlation between cholesterol and bank balance without the age influence, we could take persons from the same age group and thus control age, but if this is not possible we can statistically control the age factor and thus remove its influence on both cholesterol and bank balance. This if done is called partial correlation. That is, we can use partial and part correlation for doing the same. Sometimes in psychology we have certain factors which are influenced by large number of variables. For instance academic achievement will be affected by intelligence, work habit, extra coaching, socio economic status, etc. To find out the correlation between academic achievement with various other factors ad mentioned above can be done by Multiple Correlation. In this unit we will be learning about partial, part and multiple correlation.

3.1 OBJECTIVES

After completing this unit, you will be able to:

- Describe and explain concept of partial correlation;

- Explain, the difference between partial and semipartial correlation;
- Describe and explain concept of multiple correlation;
- Compute and interpret partial and semipartial correlations;
- Test the significance and apply the correlation to the real data;
- Compute and interpret multiple correlation; and
- Apply the correlation techniques to the real data.

3.2 PARTIAL CORRELATION (r_P)

Two variables, A and B, are closely related. The correlation between them is partialled out, or controlled for the influence of one or more variables is called as partial correlation. So when it is assumed that some other variable is influencing the correlation between A and B, then the influence of this variable(s) is partialled out for both A and B. Hence it can be considered as a correlation between two sets of residuals. Here we discuss a simple case of correlation between A and B is partialled out for C. This can be represented as $r_{AB.C}$ which is read as correlation between A and B partialled out for C. the correlation between A and B can be partialled out for more variables as well.

3.2.1 Formula and Example

For example, the researcher is interested in computing the correlation between anxiety and academic achievement controlled from intelligence. Then correlation between academic achievement (A) and anxiety (B) will be controlled for Intelligence (C). This can be represented as: $r_{\text{Academic Achievement(A) Anxiety (B) . Intelligence (C)}}$. To calculate the partial correlation (r_p) we will need a data on all three variables. The computational formula is as follows:

$$r_p = r_{AB.C} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{(1 - r_{AC}^2)(1 - r_{BC}^2)}} \qquad \text{(eq. 3.1)}$$

Look at the data of academic achievement, anxiety and intelligence. Here, the academic achievement test, the anxiety scale and intelligence test is administered on ten students. The data for ten students is provided for the three variables in the table below.

Table 3.1: Data of academic achievement, anxiety and intelligence for 10 subjects

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

In order to compute the partial correlation between the academic achievement and anxiety partialled out for Intelligence, we first need to compute the Pearson's Product moment correlation coefficient between all three variables. We have already learned to compute it in the first Unit of this Block. So I do not again explain it here.

The correlation between anxiety (B) and academic achievement (A) is – 0.369.

The correlation between intelligence (C) and academic achievement (A) is 0.918.

The correlation between anxiety (B) and intelligence (C) is – 0.245.

Give the correlations, we can now calculate the partial correlation .

$$r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1-r_{AC}^2)(1-r_{BC}^2)}} = \frac{-0.369 - (0.918 \times -0.245)}{\sqrt{(1-0.918^2)(1-(-0.245^2))}} = \frac{-.1441}{0.499} = -0.375$$

(eq.3.2)

The partial correlation between the two variables, academic achievement and anxiety controlled for intelligence, is -0.375. You will realise that the correlation between academic achievement and anxiety is -0.369. Whereas, after partialling out for the effect of intelligence, the correlation between them has almost remained unchanged. While computing this correlation, the effect of intelligence on both the variables, academic achievement and anxiety, was removed.

The following figure explains the relationship between them.

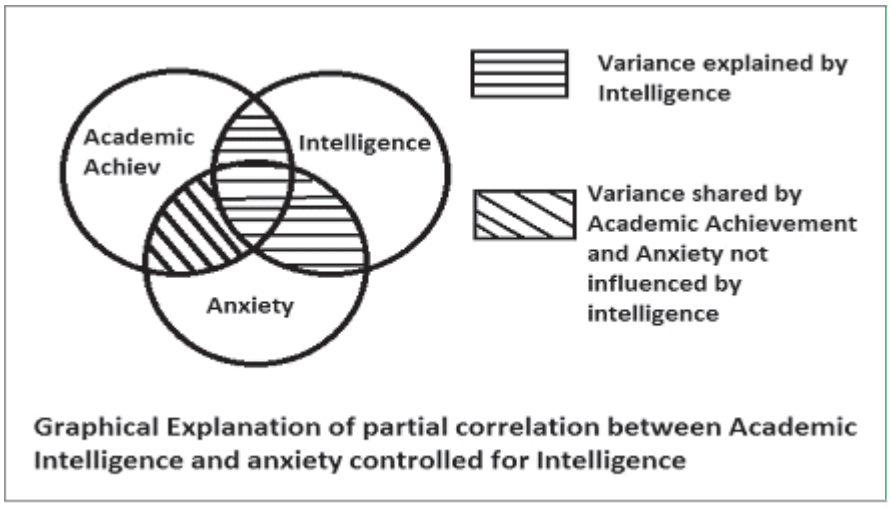

Variance explained by Intelligence
Academic Achiev
Intelligence
Anxiety
Variance shared by Academic Achievement and Anxiety not influenced by intelligence
Graphical Explanation of partial correlation between Academic Intelligence and anxiety controlled for Intelligence

Fig. 3.1: Venn diagram explaining the partial correlation

Significance testing of the partial correlation

We can test the significance of the partial correlation for the null hypothesis

$H_0 : \tilde{n}_P = 0$

and the alternative hypothesis

$H_0: \tilde{n}_P = 0$

Where, the $\tilde{n}_p$ denote the population partial correlation coefficient. The *t*-distribution is used for this purpose. Following formula is used to calculate the *t*-value.

$$t = \frac{r_P\sqrt{n-v}}{\sqrt{1-r_P^2}} \qquad \text{(eq. 3.3)}$$

Where,

r_p = partial correlation computed on sample, $r_{AB.C}$

n = sample size,

v = total number of variables employed in the analysis.

The significance of the r_p is tested at the df = n – v.

In the present example, we can employ significance testing as follows:

$$t = \frac{r_P\sqrt{n-v}}{\sqrt{1-r_P^2}} = \frac{-.375\sqrt{10-3}}{\sqrt{1-(-.375^2)}} = \frac{-0.992}{0.927} = 1.69$$

We test the significance of this value at the df = 7 in the table for t-distribution in the appendix. You will realise that at the df = 7, the table provides the critical value of 2.36 at 0.05 level of significance. The obtained value of 1.69 is smaller than this value. So we accept the null hypothesis stating that $H_0 : \tilde{n}_P = 0$.

Large sample example:

Now we take a relatively large sample example. A counseling psychologist is interested in understanding the relationship between practice of study skills and marks obtained. But she is skeptical about the effectiveness of the study skills. She believes that they can be effective because they are good cognitive techniques or they can be effective simply because the subjects believes that the study skills are going to help them. The first is attribute of the skills while second is placebo effect. She wanted to test this hypothesis. So, along with measuring the hours spent in practicing the study skills and marks obtained, she also took measures on belief that study skill training is useful. She collected the data on 100 students. The obtained correlations are as follows.

The correlation between practice of study skills (A) and unit test marks (B) is 0.69

The correlation between practice of study skills (A) and belief about usefulness of study skills (C) is 0.46

The correlation between marks in unit test (B) and belief about usefulness of study skills (C) is 0.39

$$r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1-r_{AC}^2)(1-r_{BC}^2)}} = \frac{0.69-(0.46\times 0.39)}{\sqrt{(1-0.46^2)(1-0.39^2))}} = \frac{.51}{0.82} = 0.625$$

The partial correlation between practice of study skills (A) and unit test marks (B) is 0.625. Let's test the null hypothesis about the partial correlation for a null hypothesis which states that $H_0 : \tilde{n}_P = 0$.

$$t = \frac{r_P\sqrt{n-v}}{\sqrt{1-r_P^2}} = \frac{.625\sqrt{100-3}}{\sqrt{1-.625^2}} = \frac{1.65}{0.781} = 2.12$$

The t value is significant at 0.05 level. So we reject the null hypothesis and accept that there is a partial correlation between A and B. This means that the partial correlation between practice of study skills (A) and unit test marks (B) is non-zero at population. We can conclude that the correlation between practice of study skills (A) and unit test marks (B) still exists even after controlled for the belief in the usefulness of the study skills. So the skepticism of our researcher is unwarranted.

3.2.2 Alternative Use of Partial Correlation

Suppose you have one variable which is dichotomous. These variables take two values. Some examples are, male and female, experimental and control group, patients and normal, Indians and Americans, etc. Now these two groups were measured on two variables, X and Y. You want to correlate these two variables. But you are also interested in testing whether these groups influence the correlation between the two variables. This can be done by using partial correlations. Look at the following data. This data is for male and female subjects on two variables, neuroticism and intolerance to ambiguity.

Table 3.2: Table showing gender wise data for IOA and N.

Male		**Female**	
IOA	N	IOA	N
12	22	27	20
17	28	25	15
7	24	20	18
12	32	19	12
14	30	26	18
11	27	23	13
13	29	24	20
10	17	22	9
21	34	21	19

If you compute the correlation between Intolerance of Ambiguity and neuroticism for the entire sample of male and female for 20 subjects. It is – 0.462. This is against the expectation.

This is a surprising finding which states that the as the neuroticism increases the intolerance to ambiguous situations decreases. What might be the reason for such correlation? If we examine the mean of these two variables across gender, then you will realise that the trend of mean is reversed.

If you calculate the Pearson's correlations separately for each gender, then they are well in the expected line (0.64 for males and 0.41 for females).

The partial correlations can help us in order to solve this problem. Here, we calculate the Pearson's product moment correlation between IOA and N partialled out for sex. This will be the correlation between neuroticism and intolerance of ambiguity from which the influence of sex is removed.

$$r_{AB.C} = \frac{r_{AB} - r_{AC}r_{BC}}{\sqrt{(1-r_{AC}^2)(1-r_{BC}^2)}} = \frac{-0.462 - (0.837 \times -0.782)}{\sqrt{(1-0.837^2)(1-(-0.782^2))}} = \frac{.193}{0.341} = 0.566$$

The correlation partialled out for sex is 0.57. Let's test the significance of this correlation.

$$t = \frac{r_P\sqrt{n-v}}{\sqrt{1-r_P^2}} = \frac{.566\sqrt{18-3}}{\sqrt{1-.566^2}} = \frac{2.194}{0.824} = 2.66$$

The tabled value form the appendix at *df* = 15 for 0.05 level is 2.13 and for 0.01 level is 2.95. The obtained t-value is significant at 0.05 level. So we reject the null

hypothesis which stated that population partial correlation, between IOA and N partialled out for sex is zero.

Partial correlation as Pearson's Correlation between Errors

Partial Correlation can also be understood as a Pearson's correlation between two errors.

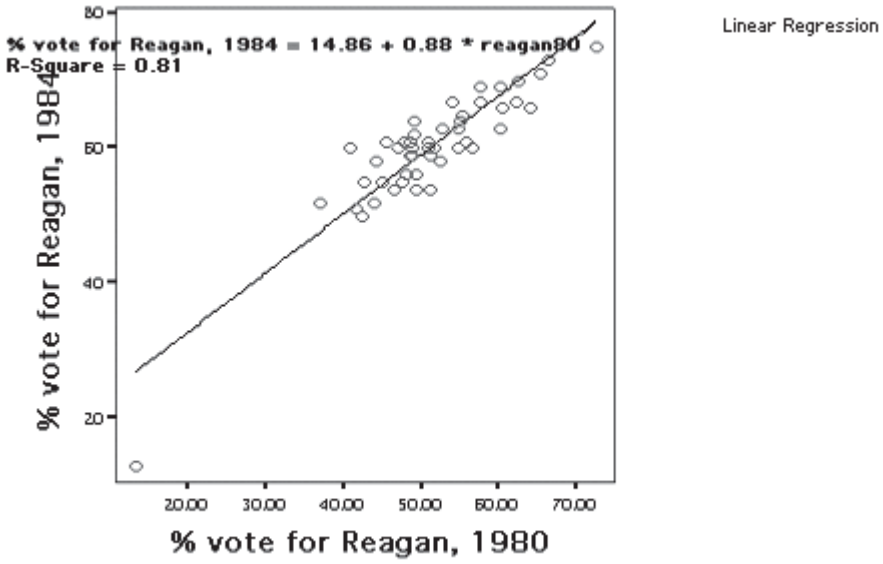
Before you proceed you need to know what is regression equation

3.3 LINEAR REGRESSION

Regression goes one step beyond correlation in identifying the relationship between two variables. It creates an equation so that values can be predicted within the range framed by the data. That is if you know X you can predict Y and if you know Y you can predict X. This is done by an equation called regression equation.

When we have a scatter plot you have learnt that the correlation between X and Y are scattered in the graph and we can draw a straight line covering the entire data. This line is called the regression line.

Here is the line and the regression equation superimposed on the scatterplot**:**



Source: http://janda.org/c10/Lectures/topic04/L25-Modeling.htm

From this line, you can predict X from Y that is % votes in1984 if known, you can find out the % of votes in 1980. Similarly if you know % of votes in 1980 you can know % of votes in 1984.

The regression line seen in the above diagram is close to the scatterplots. That is the predicted values need to be as close as possible to the data. Such a line is called the best fitting line or Regression line. There are certain guidelines for regression lines:

1) Use regression lines when there is a significant correlation to predict values.

2) Do not use if there is not a significant correlation.

3) Stay within the range of the data. For example, if the data is from 10 to 60, do not predict a value for 400.

4) Do not make predictions for a population based on another population's regression line.

The *y* variable is often termed the **criterion variable** and the *x* variable the **predictor variable**. The slope is often called the **regression coefficient** and the intercept the **regression constant**. The slope can also be expressed compactly as $ß_1 = r \times s_y/s_x$.

Normally we then predict values for *y* based on values of *x*. This still does not mean that *y* is caused by *x*. It is still imperative for the researcher to understand the variables under study and the context they operate under before making such an interpretation. Of course, simple algebra also allows one to calculate *x* values for a given value of *y*.

To obtain regression equation we use the following equation:

$$\beta = \{N * \sum xy\} - \{\sum y^2 * \sum y\} / \{(N * \sum x^2) - (\sum y^2)\}$$

Regression equation can also be written including error component 'å'

The regression equation can be written as

$$Y = \alpha + \beta X + \varepsilon \qquad \text{(eq. 4.8)}$$

Where,

Y = dependent variable or criterion variable

á = the population parameter for the y-intercept of the regression line, or regression coefficient (r=óy/ óx)

â = population slope of the regression line or regression coefficient (r*óx/ óy)

å = the error in the equation or residual

The value of á and â are not known, since they are values at the level of population. The population level value is called the parameter. It is virtually impossible to calculate parameter. So we have to estimate it. The two parameters estimated are á and â. The estimator of the á is 'a' and the estimator for â is 'b'. So at the sample level equation can be written as

$$Y = a + bX + e \qquad \text{(eq. 4.9)}$$

Where,

Y = the scores on Y variable

X = scores on X variable

a = the Y-intercept of the regression line for the sample or regression constant in sample

b = the slope of the regression line or regression coefficient in sample

e = error in prediction of the scores on Y variable, or residual

Let us take an example and demonstrate

Example: Write the regression line for the following points:

x	*y*
1	4
3	2
4	1
5	0
8	0

Solution 1: $\sum x = 21; \sum y = 7; \sum x^2 = 115; \sum y^2 = 21; \sum xy = 14$

Thus $ß_0 = [7*115 – 21*14] \div [5 * 115 - 21^2] = 511 \div 134 = 3.81$ and $ß_1 = [5*14 – 21*7] \div [5 * 115 - 21^2] = -77 \div 134 = -0.575$.

Thus the regression equation for this example is y = -0.575*x* + 3.81.

Thus if you have x , then you can find or predict y.

If you have y you can predict x.

Let's continue with the first example.

It was relationship between anxiety and academic achievement. This relationship was controlled for (partialled out for) intelligence.

In this case we can write two linear regression equations and solve them by using ordinary least-squares (OLS). They are as follows:

Academic Achievement = $a_1 + b_1 \times$ Intelligence + e_1

Where, 'a_1' is a y intercept of the regression line;

'b_1' is the slope of the line;

'e_1' is the error in the prediction of academic achievement using intelligence.

Anxiety = $a_2 + b_2 \times$ Intelligence + $å_2$

Where, 'a_2' is a y intercept of the regression line;

'b_2' is the slope of the line;

'e_2' is the error in the prediction of academic achievement using intelligence.

Now we have e_1 and e_2. They are residuals of each of the variables after intelligence explain variation in them. Meaning, e_1 is the remaining variance in academic achievement once the variance accounted for intelligence is removed. Similarly, e_2 is the variance left in the anxiety once the variance accounted for the intelligence is removed.

Now, the partial correlation can be defined as the Pearson's correlation between e_1 and e_2.

$$r_p = e_1 e_2 \qquad \text{(eq. 3.4)}$$

You will realise that this correlation is the correlation of academic achievement and anxiety, from which a linear influence of intelligence has been removed. That is called as partial correlation.

3.4 PART CORRELATION (SEMIPARTIAL CORRELATION) r_{SP}

The Part correlation is also known as semi-partial correlation (r_{sp}). Semipartial correlation or part correlation are correlation between two variables, one of which is partialled for a third variable.

In partial correlations ($r_p = r_{AB.C}$) the effect of the third variable (C) is partialled out from BOTH the variables (A and B).

In semipartial correlations ($r_{sp} = r_{A(B.C)}$), as the name suggests, the effect of third variable (C) was partialled out from only one variable (B) and NOT from both the variables.

Let's continue with the earlier example. The example was about the correlation between anxiety (A) and academic achievement (B).

In the earlier example of partial correlation, we have partialled the effect of intelligence (C) from both academic achievement and anxiety.

One may argue that the academic achievement is r=the only variable that relates to intelligence.

So we need to partial out the effect of the intelligence only from academic achievement and not from anxiety.

Now, we correlate anxiety (A) as one variable and academic achievement partialled for intelligence (B.C) as another variable.

If we correlate these two then, the correlation of anxiety (A) with academic achievement partialled for intelligence (B.C) is called as semipartial correlation ($r_{A(B.C)}$).

In fact, if there are three variables, then total six semipartial correlations can be computed. They are $r_{A(B.C)}$, $r_{A(C.B)}$, $r_{B(A.C)}$, $r_{B(C.A)}$, $r_{C(A.B)}$, and $r_{C(B.A)}$.

Formula:

In order to compute the semipartial correlation coefficient, following formula can be used.

$$r_{SP} = r_{A(B.C)} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{1 - r_{BC}^2}} \qquad \text{(eq. 3.5)}$$

Where,

$r_{A(B.C)}$ is a semipartial correlation of A with the B after linear relationship that C has with B is removed

r_{AB} Pearson's product moment correlation between A and B

r_{AC} Pearson's product moment correlation between A and C

r_{BC} Pearson's product moment correlation between B and C

Example:

Let's take the data from the earlier example of academic achievement, anxiety and intelligence. The data table 3.1 is as follows.

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

The correlation between anxiety (A) and academic achievement (B) is – 0.369.

The correlation between intelligence (C) and academic achievement (B) is 0.918.

The correlation between anxiety (A) and intelligence (C) is – 0.245.

Given the correlations, we can now calculate the semipartial correlation (r_{SP}) as follows. We are not computing the correlation coefficients, simply because you have already learned to compute the correlations earlier. The formula for semipartial correlation is as follows:

$$r_{SP} = r_{A(B.C)} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{1 - r_{BC}^2}} \qquad \text{(eq. 3.6)}$$

Now we can calculate semipartial correlation by using this formula.

$$r_{AB.C} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{1 - r_{BC}^2}} = \frac{-0.369 - (-0.245 \times 0.918)}{\sqrt{1 - (0.918^2)}} = \frac{-.1441}{0.3966} = -0.363$$

The semipartial correlation between anxiety and academic achievement after the linear relationship between the academic achievement and intelligence is removed is -0.363.

The significance of the semipartial correlation can be tested by using t-distribution. The null hpothsis and the alternate hypothesis are as follows.

H_0: $\tilde{n}_{SP} = 0$

H_A: $\tilde{n}_{SP} \neq 0$

Where, the $\tilde{n}_{SP}$ is the semipartial correlation in the population. We test the null hypothesis whether the semipartial correlation in the population is zero. This can be done by using following formula

$$t = \frac{r_{SP}\sqrt{n - v}}{\sqrt{1 - r_{SP}^2}} \qquad \text{(eq. 3.7)}$$

Where,

t = students t-value

r_{SP} = semipartial correlation computed on sample,

n = sample size,

v = number of variables used in the analysis

The significance of this t-value is tested at the df = n – v. when three variables are involved then the df is n – 3.

For our example, the t-values can be computed as follows:

$$t = \frac{-0.363\sqrt{10-3}}{\sqrt{1-(-0.363^2)}} = -1.032$$

The obtained t-value is tested at df = n – v = 10 – 3 = 7.

The t-value at .05 level is 2.364. The obtained t-value is smaller than that. So we accept the null hypothesis that the population semipartial correlation is zero.

It has an interesting implication for our data. The correlation between anxiety and academic achievement is zero in the population if the linear relationship between academic achievement and intelligence is removed.

3.4.1 Semipartial Correlation: Alternative Understanding

Partial Correlation can also be understood as a Pearson's correlation between a variable and error (residual).

Let us continue with the first example. It was relationship between anxiety and academic achievement. This relationship was controlled for (partialled out for) intelligence and academic achievement. (So far you have not learned regression and you may not follow some of the points and equations. So you can revisit this discussion after learning regression.)

In this case we can write a linear regression equation and solve them by using ordinary least-squares (OLS). They are as follows:

Academic Achievement = $a_1 + b_1 \times$ Intelligence + e_1

Where, 'a_1' is a y intercept of the regression line;

'b_1' is the slope of the line;

'e_1' is the error in the prediction of academic achievement using intelligence.

Now we have e_1. It is a residuals of academic achievement after intelligence explain variation in academic achievement.

That is, e_1 is the remaining variance in academic achievement once the variance accounted for intelligence is removed.

Now, the semipartial correlation can be defined as the Pearson's correlation between anxiety and e_1.

You will realise that this correlation is the correlation of academic achievement and anxiety, from which a linear influence of intelligence on academic achievement has been removed.

That is called the semipartial correlation.

(Since you have not learned the regression equation you may not be able to understand this point. So revisit this point after learning regression.)

3.5 MULTIPLE CORRELATION COEFFICIENT (R)

The multiple correlation coefficient denoting a correlation of one variable with multiple other variables. The multiple correlation coefficient is denoted as $R_{A.BCD\ldots k}$ which denotes that A is correlated with B, C, D, up to *k* variables.

For example, we want to compute multiple correlation between A with B and C then it is expressed as $R_{A.BC}$. In this case we create a linear combination of the B and C which is correlated with A.

We continue with the same example which we have discussed for partial and semipartial correlations. This example has academic achievement, anxiety and intelligence as three variables. The correlation between academic achievement with the linear combination of anxiety and intelligence is multiple correlation.

This denotes the proportion of variance in academic achievement explained by intelligence and anxiety. We denote this as

$R_{\text{(Academic Achievement. Intelligence, Anxiety),}}$ which is a multiple correlation.

Often, it is used in the context of regression, where academic achievement is a criterion variable and intelligence and anxiety are called as predictors.

We are not using regression equation since you have not learned it. The Multiple R can be calculated for two predictor variable as follows.

$$R_{\text{A . BC}} = \sqrt{\frac{r_{AB}^2 + r_{AC}^2 - 2r_{AB}r_{AC}r_{BC}}{1 - r_{BC}^2}} \qquad \text{(eq. 3.7)}$$

Where,

$R_{\text{A . BC}}$ = is multiple correlation between A and linear combination of B and C.

r_{AB} = is correlation between A and B

r_{AC} = is correlation between A and C

r_{BC} = is correlation between B and C

Example

We shall continue with the earlier data.

The data table 3.1 is as follows.

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

The correlation between anxiety (A) and academic achievement (B) is – 0.369.

The correlation between intelligence (C) and academic achievement (B) is 0.918.

The correlation between anxiety (A) and intelligence (C) is – 0.245.

The multiple correlation can be calculated as follows.

$$R_{\text{A . BC}} = \sqrt{\frac{r_{AB}^2 + r_{AC}^2 - 2r_{AB}r_{AC}r_{BC}}{1 - r_{BC}^2}}$$

$$= \sqrt{\frac{-0.369^2 + 0.918^2 - 2 \times -.369 \times .918 \times -.245}{1 - (-0.245)^2}}$$

$$= \sqrt{\frac{0.813}{0.94}}$$

$$= 0.929$$

This means that the multiple correlation between academic achievement and the linear combination of intelligence and anxiety is 0.929 or 0.93. We have earlier learned that the square of the correlation coefficient can be understood as percentage of variance explained.

The R^2 is then percentage of variance in academic achievement explained by the linear combination of intelligence and anxiety. In this example the R^2 is 0.929^2 which is 0.865. The linear combination of intelligence and anxiety explain 86.5 percent variance in the academic achievement.

We have already converted the R into the R^2 value. The R^2 is the value obtained on a sample. The population value of the R^2 is denoted as P^2. The R^2 is an estimator of the P^2.

But there is a problem in estimating the P^2 value from the R^2 value.

The R^2 is not an unbiased estimator of the P^2.

So we need to adjust the value of the R^2 in order to make it unbiased estimator.

Following formula is used for this purpose.

Let $\tilde{R}^2$ denote an adjusted R^2. Then $\tilde{R}^2$ can be computed as follows:

$$\tilde{R}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \qquad \text{(eq. 3.8)}$$

Where,

$\tilde{R}^2$ = adjusted value of R^2

k = number of predicted variables (or the variable for which a linear combination is created)

n = sample size

For our example the $\tilde{R}^2$ value need to be computed.

$$\tilde{R}^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

$$\tilde{R}^2 = 1 - \frac{(1-0.865)(10-1)}{10-2-1}$$

$$\tilde{R}^2 = 1 - \frac{1.217}{7} = 0.826$$

So the unbiased estimator of the R^2 the adjusted value, $\tilde{R}^2$, is 0.826 which is smaller than the value of R^2. It is usual to get a smaller adjusted value.

The significance testing of the R:

This can be used for the purpose of the significance testing. The null hypothesis and the alternative hypothesis employed for this purpose are

$$H_0 : P^2 = 0$$
$$H_A : P^2 \neq 0$$

The null hypothesis denotes that the population R^2 is zero whereas the alternative hypothesis denotes that the population R^2 is not zero.

The *F*-distribution is used for calculating the significance of the R^2 as follows:

$$F = \frac{(n-k-1)R^2}{k(1-R^2)} \qquad \text{(eq. 3.9)}$$

The value of R^2 can be adjusted R^2 ($\tilde{R}^2$ = .825) value of the R^2 value (.865).

When the sample size is small, it is recommended that $\tilde{R}^2$ value be used. As the sample size increase the difference between the resulting F values reduce considerably. Since our sample is obviously small, we will use unbiased estimator.

$$F = \frac{(n-k-1)R^2}{k(1-R^2)}$$

$$F = \frac{(10-2-1)\times 0.826}{2\times(1-0.826)} = \frac{5.783}{0.348} = 16.635$$

The degrees of freedom employed in the significance testing of this F value are $df_{num} = k$ and $df_{denominator} = n - k - 1$.

For our example the degrees of freedom are as follows:

$df_{num} = k = 2$

$df_{denominator} = n - k - 1 = 10 - 2 - 1 = 7$

The tabled value for the $F_{(2, 7)} = 4.737$ at 0.05 level of significance and the $F_{(2, 7)} = 9.547$ at 0.01 level of significance. The calculated value of the F 16.635 is greater than the critical value of F. so we reject the null hypothesis and accept the alternative hypothesis which stated that the P^2 is not zero at less than 0.01 level.

You could have computed the F value using the R^2 instead of adjusted R^2 ($\tilde{R}^2$).

Often the statistical packages report the significance of R^2 and not significance of adjusted R^2 ($\tilde{R}^2$).

It is the judgment of the researcher to use either of them. In the same example if R^2 value is substituted for the adjusted R^2 ($\tilde{R}^2$) value then the F is 22.387 that is significant at .01 level.

3.6 LET US SUM UP

In this unit we have learned about the interesting procedures of computing the correlations. Especially, when we are interested in controlling for one or more variable. The multiple correlations provide us with an opportunity to calculate correlations between a variable and a linear combination of other variable. You practice them by solving some of the example given below, and you will understand the use of them.

3.7 UNIT END QUESTIONS

Problems

1) A clinical psychologist was interested in testing the relationship between health and stress. But she realised that coping skills will have an influence on this relationship so she administered General Health Questionnaire, Stress Scale and a coping scale. The data was collected on 15 individuals. Calculate the multiple correlation for this problem and test the significance. The data is as follows:

Health	Stress	Coping
9	5	18
10	5	21
8	7	17
7	4	16
8	7	22
9	6	19
12	8	25
11	3	17
10	5	20
14	6	22
7	7	18
9	9	22
6	7	17
16	3	20
14	8	26

In addition, answer the following questions:

Which correlation coefficient she should compute, if she wants to control the relationship between stress and health for the effect of coping?

Write a null and alternative hypothesis for partial correlation.

Calculate the partial correlation between stress and health controlled for the effect of coping.

Test the significance of the relationship.

Which correlation coefficient she should compute, if she wants to control the relationship between stress and health for the effect of coping only on stress?

Write a null and alternative hypothesis for part correlation.

Calculate the part correlation between stress and health controlled for the effect of coping on stress and test the significance.

Which correlation coefficient she should compute, if she wants to know the relationship between health (Y) and linear combination of stress (X_1) and coping (X_2)?

Answer:

a) Partial correlation; b) H_0: $r_p = 0$ and H_A : $r_p \neq 0$; c) –.77; d) significant at 0.01 level; e) part (semipartial); f) H_0: $r_{sp} = 0$ and H_A : $r_{sp} \neq 0$; g) –.62, $p < .05$. h) Multiple correlation; i) $R = 0.86$, $p < .01$.

2) A social psychologist was interested in testing the relationship between attitude towards women (ATW) and openness to values (OV). But she realised education (EDU) will influence this relationship so she administered attitude to women scale, openness to values scale and also recorded the years spent in formal education. The data was collected on 10 individuals.

Calculate the multiple correlation for this problem and test the significance.

The data is as follows:

ATW	OV	Edu
2	7	14
4	10	13
8	14	11
7	13	9
8	9	5
9	10	14
1	6	5
0	9	6
6	12	11
5	10	12

In addition, answer the following questions:

1) Which correlation coefficient she should compute, if she wants to control the relationship between ATW and OV for the effect of EDU?

2) Write a null and alternative hypothesis for partial correlation.

3) Calculate the partial correlation between ATW and OV controlled for the effect of EDU.

4) Test the significance of the relationship.

5) Which correlation coefficient she should compute, if she wants to control the relationship between ATW and OV for the effect of EDU only on OV?

6) Write a null and alternative hypothesis for part correlation.

7) Calculate the part correlation between ATW and OV controlled for the effect of EDU on OV and test the significance.

8) Which correlation coefficient she should compute, if she wants to know the relationship between ATW (Y) and linear combination of OV (X_1) and EDU (X_2)?

9) Write a null and alternative hypothesis for multiple correlation.

Answer:

a) Partial correlation; b) H_0: $r_p = 0$ and H_A : $r_p \neq 0$; c) .64; d) insignificant (the n is small); e) part (semipartial); f) H_0: $r_{sp} = 0$ and H_A : $r_{sp} \neq 0$; g) 0.61, p >.05. h) Multiple correlation; i) R = 0.67, p >.05.

3.8 SUGGESTED READINGS

Garrett, H.E. (19). *Statistics In Psychology And Education.* Goyal Publishing House, New Delhi.

Guilford, J.P.(1956). *Fundamental Statistics in Psychology and Education.* McGraw Hill Book company Inc. New York.

UNIT 4 BIVARIATE AND MULTIPLE REGRESSION

Structure

4.0 Introduction

4.1 Objectives

4.2 Bivariate and Multiple Regression

- 4.2.1 Predicting one Variable from Another
- 4.2.2 Plotting the Relationship
- 4.2.3 Mean, Variance and Covariance: Building Blocks of Regression
- 4.2.4 The Regression Equation
- 4.2.5 Ordinary Least Squares (OLS)
- 4.2.6 Significance of Testing of b.
- 4.2.7 Accuracy of Prediction
- 4.2.8 Assumptions Underlying Regression
- 4.2.9 Interpretation of Regression

4.3 Standardised Regression Analysis and Standardised Coefficients

4.4 Multiple Regression

4.5 Let Us Sum Up

4.6 Unit End Questions

4.7 Suggested Readings

4.0 INTRODUCTION

Psychologists, as other scientists, are also interested in prediction. Since our domain of enquiry relates with human behaviour, our predictions are associated with human behaviour. We are interested in knowing how human beings will behave provided we have some information about them. It is not that we all the time depend on theories such as psychoanalysis, behaviourism or cognitive in order to predict human behaviour. There are also statistical methods which can help predict certain phenomenon of human behaviour. We would study in this unit the statistical methods that can be used for the purpose of prediction. These statistical methods are called Regression. We will first learn the concept of regression, then learn how to plot the relationship between variables, and learn to work out The Regression Equation. We will also deal with how far we can be accurate in predicting with the help of regression equation by the help of tests of significance. Finally we will be dealing with how to interpret regression and deal with also Multiple regression, that is, which variables influence a particular phenomenon.

4.1 OBJECTIVES

After completing this unit, you will be able to:

- Describe and explain concept of regression correlation;
- Explain, describe and differentiate between bivariate regression and multiple regresion;

- Describe and explain concept of multiple correlation;
- Develop a regression equation;
- Compute the a and b of bivariate regression by using OLS;
- Test the significance of regression;
- Interpret regression results;
- Apply the regression techniques to the real data;
- Explain Multiple regression; and
- Use Multiple regression in real data.

4.2 BIVARIATE AND MULTIPLE REGRESSION

We always see that the meteorology department predicts the rain, the economists predict the outcome of a particular policy, financial experts predict the share market, the election experts predict the outcome of voting and so on. Similarly using statistical method, psychologists too can predict certain human behaviours. For example, one can predict the examination marks after writing the examination by checking what questions we have attempted and how we have faired in it and what marks we can expect for each question and so on.

Most of us are interested in this exercise of prediction. On many occasions, we also predict the behaviour of our friends, colleagues and family. Predicting and trying to speculate about what might happen in future is integral part of human curiosity. While there are many theories of psychology and personality that would help us predict behaviours, one can also predict a certain phenomenon in terms of statistics. This method is called regression in statistics.

The simplest form of the regression is simple linear regression (at times also called as bivariate regression). Carl Frederick Gauss discovered a method of least squares (1809) and later on developed Gauss-Markov theorem (1821). Sir Fransis Galton contributed to the method of regression and also gave the name.

Let us see what this is all about. Let us say we have data on two variables (Y and X), and we create an equation, called regression equation, which later on helps us in predicting the score of one variable (Y) by simply using the scores on another variable (X). Let us learn about the utility of the regression analysis, how to do it, how to test the significance and issues surrounding it.

Regression analysis tries to predict Y variable from X variable. In the general form, it tries to predict Y from a X_1, X_2, …, X_k, where k is number of predictor variables. Initially we will learn about two variable prediction, one of which is a predictor and the other one will be predicted. Then we will look at the general form of Regression.

Just think of the variables that can be used in prediction in psychology. Look at the following statements. (See the box below)

- Stress leads to health deterioration.
- Openness increases creativity.
- Extraversion increases social acceptance.

- Social support influences coping with mental health problems.
- Stigma about mental illness decides the help seeking behaviour.
- Parental intelligence leads to child's intelligence
- Attitude to job and attrition depends on affective commitment to the organisation.

What do you see in common in all these statements?

All the statements above have two variables.

One of the variables can potentially predict the other variable.

4.2.1 Predicting One Variable from Another

Let us now consider the problem of prediction. How to predict Y from X.

The Y variable is called the dependent variable. It is also called as criterion variable. It is the variable that has to be predicted.

The X variable is a called as independent variable. It is also called as predictor variable (Please note that in experimental psychology we define independent variable as the variable that is manipulated by the experimenter, whereas in regression the term is used less strictly. In Regression, the independent variable is not manipulated by the researcher or experimenter.)

If X is predicting Y, then typically it is said that 'X is regressed on Y'.

Let's identify the X and Y in our statements given in the box.

In the first statement Stress (X) lead to the health (Y) deterioration

In the second statement, Openness (X) increases the creativity (Y).

In the third statement, Extroversion (X) increases the social acceptance (Y).

In fourth statement, Social support (X) influences the coping with mental health problems (Y).

In the fifth statement, Stigma about mental illness (X) decided the help seeking behaviour (Y).

In the sixth statement, Parental intelligence (X) leads to child's intelligence (Y).

In the last statement, Attitude (Y) to job and attrition depends on affective commitment(X) to the organisation

Before we learn how to do the regression, we shall quickly browse through the basic concepts in regression analysis.

4.2.2 Plotting the Relationship

We have already learned to plot the scatter plots. We shall try to plot a scatter and try to understand regression graphically.

The perfect relationship

Look at the following example. You have data of five swimmers on two variables, hours of practice per day (X) and time taken (Y).

Swimmer	hours of practice per day	Time taken (in seconds)
A	1	50
B	2	45
C	3	40
D	4	35
E	5	30

Now plot the relationship between them as a scatter. You know how to do that. We have now tried to draw a line that passes through all the data points in the scatter. And we have successfully done it.

Looking at figure 4.1 you realise that as the number of hours spent in practice increase the time taken is reducing. There is a perfect linear relationship between them. This means that you can draw a line on the scatter that passes through all the data points on the scatter.

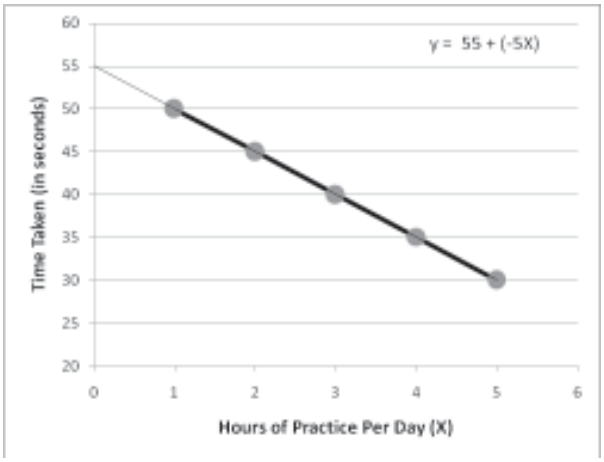


Fig. 4.1: Figure showing the data between number of hours spent in practice and time take.

For this data, the slope of the line can be calculated by using a simple technique.

$$Slope = \frac{Y_2 - Y_1}{X_2 - X_1} \qquad \text{(eq. 4.1)}$$

Where Y_2 and Y_1 are any two points on Y axis and X_2 and X_1 are corresponding two points on X axis.

For example, take $Y_2 = 45$ and $Y_1 = 40$ and corresponding X_2 and X_1 are 2 and 3. The slope is

$$Slope = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{45 - 40}{2 - 3} = \frac{5}{-1} = -5 \qquad \text{(eq. 4.2)}$$

The slope of the line is – 5.

The point at which the line passes through the Y axis (the Y intercept of the line) is 55.

Now, if we ask about the unknown score, 6 hours of practice per day, then the predicted X score is 25 seconds (which is very close the world record).

How have we obtained it? we have solved it for a equation of straight line. That equation is

$$Y = a + bX \qquad \text{(eq. 4.3)}$$

Where a = point where the line passes the Y axis and

b = is a slope of the line.

We have $a = 55$ and $b = -5$. So for X = 6 the Y will be

(eq. 4.4)

The Imperfect Relationship.

But the problem is the real data will not be so systematic and all data points in scatter will not fall on a straight line.

Look at the following example of the stigma and visits to mental health professionals. The Table 4.2 shown below display the data of stigma and number of appointments missed to mental health professional.

Table 4.2: Data of stigma and number of appointments missed to mental health professional

Patient	Stigma scores	Number of appointments missed
1	60	5
2	50	2
3	70	9
4	73	6
5	64	9
6	68	4
7	56	3
8	54	8
9	49	3
10	66	11

This data was obtained from ten patients who are suffering due to mental illness. The data was collected on King, Show and others (2007) Stigma scale and the data were obtained on number of visits missed by the patients. The data is plotted in the scatter plot below.

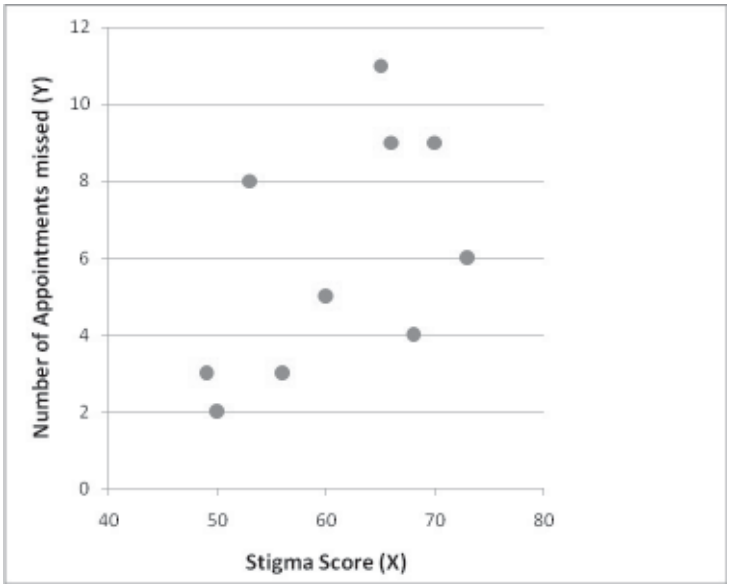


Fig. 4.2: Scatter showing the relationship between stigma and number of appointments missed

Now you will realise that it is not possible to draw a straight line that passes through all the data points. Then how to know the relationship between X and Y and then predict the scores of Y from scores of X. How to draw the straight line for this data? This is a problem one would face with real data. The linear regression analysis solves this problem.

4.2.3 Mean, Variance and Covariance: Building Blocks of Regression

In order to understand the building blocks of regression we must describe some of the terms such as (i) the mean, (ii) variance, and (iii) covariance which are presented in the following section.

i) Mean

Mean of variable X (symbolised as $\bar{X}$) is sum of scores ($\sum_{i=1}^{n} X_i$) divided by number of observations (n). The mean is calculated in following way.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \text{(eq. 4.5)}$$

You have learned this in the first block. We will need to use this as a basic element to compute correlation.

ii) Variance

The variance of a variable X (symbolised as S_X^2) is the sum of squares of the deviations of each X score from the mean of X ($\sum (X - \bar{X})^2$) divided by number of observations (n).

$$^{2}_{X} = \frac{\sum (X - \bar{X})^2}{n} \qquad \text{(eq. 4.6)}$$

You have already learned that standard deviation of variable X , symbolised as S_X , is square root of variance of X, symbolised as .

iii) Covariance

The covariance between X and Y (Cov_{XY} or S_{XY}) can be stated as

$$Cov_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n} \qquad \text{(eq. 4.7)}$$

Covariance is a number that indicates the association between two variables. To compute covariance, deviation of each score on X from its mean ($\bar{X}$) and deviation of each score on Y from its mean ($\bar{Y}$) is initially calculated.

Then products of these deviations are obtained.

Then, these products are summated.

This sum gives us the numerator for covariance.

Divide this sum by number of observations (*n*).

The resulting number is covariance.

4.2.4 The Regression Equation

The regression equation can be written as

$$Y = \alpha + \beta X + \varepsilon \quad \text{(eq. 4.8)}$$

Where,

Y = dependent variable or criterion variable

á = the population parameter for the y-intercept of the regression line, or regression coefficient (r=óy/ óx)

â = population slope of the regression line or regression coefficient (r*óx/ óy)

å = the error in the equation or residual

The value of á and â are not known, since they are values at the level of population. The population level value is called the parameter. It is virtually impossible to calculate parameter. So we have to estimate it. The two parameters estimated are á and â. The estimator of the á is 'a' and the estimator for â is 'b'. So at the sample level equation can be written as

$$Y = a + bX + e \quad \text{(eq. 4.9)}$$

Where,

Y = the scores on Y variable

X = scores on X variable

a = the Y-intercept of the regression line for the sample or regression constant in sample

b = the slope of the regression line or regression coefficient in sample

e = error in prediction of the scores on Y variable, or residual

$$\hat{Y} = a + bX \quad \text{(eq. 4.10)}$$

Where, Y = predicted value of Y in sample. This value is not an actual value but the value of Y that is predicted using the equation $\hat{Y} = a + bX$. So we can write error as by substituting the in the earlier equation.

$$S_x^2 \quad \text{(eq. 4.11)}$$

$$Y - \hat{Y} = e \quad \text{(eq. 4.12)}$$

This is a useful expression. We shall use it while computing the statistical significance of the regression and will also be useful for understanding the least squares.

4.2.5 Ordinary Least Squares (OLS)

Just recall the data between the stigma scores and number of appointments missed by the person. Now, if we have to draw the straight line that will explain the relationship between the stigma scores and number of appointments missed, then there will be many such lines possible. Out of them, which line we should consider as the best fit line?

It is not possible to draw a straight line that will pass through all the points. And many lines are possible with the earlier equation $Y = a + bX + e$

This problem is solved by the method of least squares or ordinary least squares (OLS).

One easy way to judging how good the line is, is to know how close various values of $\hat{Y}$ are to corresponding values of Y, which means to check how close predicted value ($\hat{Y}$) is to the actual value of the Y.

These predicted values are computed by using the various values of X in the data.

But how to decide what is the best fit?

One logical solution to this problem is to look at the error term, e. the e is defined as

$$Y - \hat{Y} = e$$

Which means,

$$Y - (a + bX) = e \qquad \text{(eq. 4.13)}$$

The $Y - \hat{Y}$ is the error in prediction of the Y.

This error is called as an obtained residual of the regression.

The best line is the one that minimises this residual.

Some of the predicted values of Y will be higher than the actual value of Y and some would be lower, and hence the sum of residual will be zero.

In order to take care of this problem, the summation is not done over the $Y - \hat{Y}$ Instead the $\sum\left(Y - \hat{Y}\right)^2$ is summated. An attempt to *minimise the sum of the squared errors* — minimise the $\sum\left(Y - \hat{Y}\right)^2$ is made, this is called as least squares.

Calculation of a and b

The values for *a* and *b* that minimises the sum of the squared errors — minimise the $\sum\left(Y - \hat{Y}\right)^2$ need to be calculated. The *b* can be calculated as follows.

$$b = \frac{Cov_{XY}}{S_X^2} \qquad \text{(eq. 4.14)}$$

Where,

Cov_{XY} = covariance between X and Y. This is given by the formula $\sum$ (X-X)(Y-Y) / N S_X^2 = variance of X

The *b* is covariance of X and Y divided by the variance of X. it can be rewritten as

$$b = \frac{\frac{\sum (X - \bar{X})(Y - \bar{Y})}{n}}{S_X^2} \qquad \text{(eq. 4.15)}$$

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{nS_X^2} \qquad \text{(eq. 4.16)}$$

The *a* can be calculated as follows by using our earlier equation.

$$\bar{Y} = a + b\bar{X} \qquad \text{(eq. 4.17)}$$

$$a = \bar{Y} - b\bar{X} \qquad \text{(eq. 4.18)}$$

Once we know how to calculate *a* and *b*, then we can solve the problem of regression. Let's now solve the example we have started with. The example was about the predicting the number of appointments missed by the patient (Y) by using the Stigma scale scores (X). The data is as follows:

Table 4.3: Table showing the computation of *a* and *b*.

Patient	**Stigma scores**	**Number of appointments missed**	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
1	60	5	-1	-1	1	1	1
2	50	2	-11	-4	121	16	44
3	70	9	9	3	81	9	27
4	73	6	12	0	144	0	0
5	64	9	3	3	9	9	9
6	68	4	7	-2	49	4	-14
7	56	3	-5	-3	25	9	15
8	54	8	-7	2	49	4	-14
9	49	3	-12	-3	144	9	36
10	66	11	5	5	25	25	25
X	$\sum X$ =610	$\sum Y$ =60			$\sum (X - \bar{X})^2$ = 648	$\sum (Y - \bar{Y})^2$ = 86	$\sum (X - \bar{X})(Y - \bar{Y})$ = 129
n = 10	$\bar{X} = 61$	$\bar{Y} = 6$					

$$S_X = \sqrt{\sum (X - \bar{X})^2 / n} = 8.50$$

$$S_Y = \sqrt{\sum (Y - \bar{Y})^2 / n} = 2.93$$

$$Cov_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n} = \frac{129}{10} = 12.9$$

$$b = \frac{Cov_{XY}}{S_X^2} = \frac{12.9}{8.50^2} = \frac{12.9}{64.8} = 0.1991$$

$$a = \bar{Y} - b\bar{X} = 6 - (0.1991 \times 61) = -6.144$$

Step 1. You need scores of subjects on two variables. We have scores on ten subjects on two variables, the Stigma scores (X) and number of appointments missed (Y).

Then list the pairs of scores on two variables in two columns.

The order will not make any difference.

Remember, same individuals' two scores should be kept together.

Label the predictor variable as X and criterion as Y.

Step 2. Compute the mean of variable X and variable Y. It was found to be 61 and 6 respectively.

Step 3. Compute the deviation of each X score from its mean ($\overline{X}$) and each Y score from its own mean ($\overline{Y}$). This is shown in the column labeled as $X - \overline{X}$ and $Y - \overline{Y}$. As you have learned earlier, the sum of these columns has to be zero.

Step 4. Compute the square of $X - \overline{X}$ and $Y - \overline{Y}$. This is shown in next two columns labeled as $(X - \overline{X})^2$ and $(Y - \overline{Y})^2$. Then compute the sum of these squared deviations of X and Y.

The sum of squared deviations for X is 648 and for Y it is 86.

Divide them by *n* to obtain the standard deviations for X and Y. The was found to be 8.49. Similarly, the S_y was found to be 3.09.

Step 5. Compute the cross-product of the deviations of X and Y. These cross-products are shown in the last column labeled as $(X - \overline{X})$ and $(Y - \overline{Y})$. Then obtain the sum of these cross-products. It was found to be 129. Now, we have all the elements required for computing *b*.

Step 6. Compute the covariance between X and Y, which turned out to be 12.9.

Step 7. Compute the *b* value by dividing the covariance XY (Cov_{XY}) by the variance of . We compute S_x^2 by taking *n* as a denominator the S_x^2 value is 64.8. The b is found to be 0.1991. Now we can easily compute the *a* which is

$$a = \overline{Y} - b\overline{X} = 6 - (0.1991 \times 61) = -6.144 .$$

Once the *a* and *b* are computed, we can write the regression equation to get the predicted values of Y as follows:

$$\hat{Y} = a + bX \qquad \text{(eq. 4.19)}$$

$$\hat{Y} = -6.144 + (0.1991 \times X)$$

Now we can compute the predicted values for each of the X value. For example the predicted value for the first X value (60) is as follows:

$$5.80 = -6.144 + (0.1991 \times 60)$$

In this way you can compute the predicted Y value for each of the X score. Now you realise that this value is not Y value but the predicted Y value obtained from X.

Now look at the table below. It gives the X, Y and Predicted Y values.

Table 4.4: Table showing the computation of the significance for the *b*, the slope of the line

Ss	Stigma scores (X)	Number of appointments missed (Y)	Predicted value of Y	Residual $Y-\hat{Y}=e$	Residual $(Y-\hat{Y})^2=e$	Variance explained $\hat{Y}-\overline{Y}$	Variance explained Squared $(\hat{Y}-\overline{Y})^2$
1	60	5	5.80	-0.80	0.64	-0.20	0.04
2	50	2	3.81	-1.81	3.28	-2.19	4.80
3	70	9	7.79	1.21	1.46	1.79	3.21
4	73	6	8.39	-2.39	5.71	2.39	5.71
5	64	9	6.60	2.40	5.77	0.60	0.36
6	68	4	7.39	-3.39	11.52	1.39	1.94
7	56	3	5.00	-2.00	4.02	-1.00	0.99
8	54	8	4.61	3.39	11.52	-1.39	1.94
9	49	3	3.61	-0.61	0.37	-2.39	5.71
10	66	11	7.00	4.00	16.04	1.00	0.99
Sum	610	60	60	0	60.32	0	25.68

With the availability of residual, we can obtain the sum of squared residual. The sum of squared residual is 60.32. This is the minimum value that can be obtained if a straight line is drawn for the relationship between X and Y.

There is no other line than can give value as small as this.

So this line is considered as a best fit line.

The mean of Y is 6. So we can now obtain an interesting expression. This expression is $\hat{Y}-\overline{Y}$.

This will provide us the amount of variance in Y explained by the predicted value of Y which is $\hat{Y}$.

The sum of this difference is bound to be zero. So we square the difference.

The sum of square of the difference between predicted value of Y and mean of Y is given below:

$$\sum(\hat{Y}-\overline{Y})^2 ,$$

This is the amount of variance explained in the Y by the predicted value of the Y. This can be expressed as follows:

Total Variance in Y = Variance Explained by Regresion + Residual variance

(eq. 4.20)

This can be written as

$$Y-\overline{Y}=(\hat{Y}-\overline{Y})+(Y-\hat{Y}) \quad \text{(eq. 4.21)}$$

$$\sum Y-\overline{Y}=\sum(\hat{Y}-\overline{Y})+\sum(Y-\hat{Y}) \quad \text{(eq. 4.22)}$$

Since the summation of these differences are zero, we square the difference. The equation can be rewritten as

$$\sum (Y-\bar{Y})^2 = \sum (\hat{Y}-\bar{Y})^2 + \sum (Y-\hat{Y})^2 \qquad \text{(eq. 4.23)}$$

Where,

$\sum (Y-\bar{Y})^2$ = Total variance in Y. Total sum of squares (SS_T).

$\sum (\hat{Y}-\bar{Y})^2$ = Variance in Y explained by X. Sum of squares explained ($SS_{Regression}$).

$\sum (Y-\hat{Y})^2$ = variance in Y not explained by X. Residual sum of squares ($SS_{Residual}$).

$$SS_{Total} = SS_{Regresion} + SS_{Residual} \qquad \text{(eq. 4.24)}$$

Look at the figure below. You will understand the division of SS_{Total} into $SS_{Regression}$ and $SS_{Residual}$.

It shows that the distance between the $\bar{Y}$ and Y is total deviation of that Y value from $\bar{Y}$. This is shown as $(Y-\bar{Y})$.

From this total deviation or variation, the explained variation is distance between $\bar{Y}$ and the predicted Y value. This is shown as $\hat{Y}-\bar{Y}$.

This is explained by the regression line. The distance that regression equation fails to explain is between Y and predicted value of Y. This distance is residual or remaining variance that regression equation cannot explain. This is shown as $Y-\hat{Y}$.

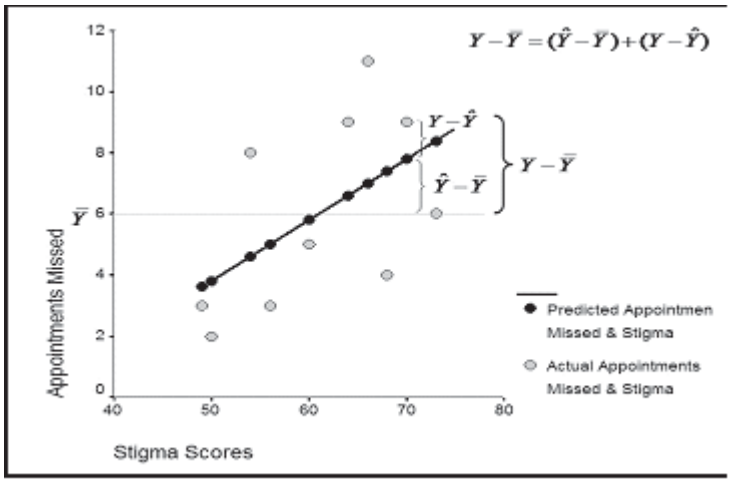


Fig. 4.3: The figure showing the scatter of X and Y, the regression line, and also explains the variance explained, residual and total.

4.2.6 Significance Testing of b

The *F*-distribution is employed to test the significance of the *b*.

The slope or regression coefficient obtained on sample is an estimator of the population slope or population regression coefficient called as â.

We already have completed the basics for computing the *F*-distribution.

$$F = \frac{S^2_{Between}}{S^2_{Within}} \qquad \text{(eq. 4.25)}$$

In case of regression, the same formula is used. The sum of squares total, sum of squares regression, and sum of squares residual have already been computed. We will use them now. Look at the table below.

Table 4.5: Table showing the computation of significance of *b*.

Source	Sum of Squares	*df*	S^2	F
Regression	$\sum(\hat{Y}-\bar{Y})^2$	k	$\frac{\sum(\hat{Y}-\bar{Y})^2}{k}$	$\frac{S^2_{Regresion}}{S^2_{Residual}}$
Residual	$\sum(Y-\hat{Y})^2$	$n-k-1$	$\frac{\sum(Y-\hat{Y})^2}{n-k-1}$	
Total	$\sum(Y-\bar{Y})^2$	$n-1$		

Where, n = sample size, and k = number of independent variables.

The null and the alternative hypothesis tested are as follows:

The F is computed for our example.

Table 4.6: Table showing the computation of the *F*-statistics for the data.

Source	Sum of Squares	*df*	S^2	F
Regression	25.68	1	25.68	$\frac{25.68}{7.54}=3.41$
Residual	60.32	8	7.54	
Total	86	9		

The *F*-value needs to be tested for its significance. The *F*-value at numerator *df* = 1, and denominator *df* = 8 at 0.05 level is 5.31. The obtained value of the *F* is smaller than the tabled value of the F. This means that we need to accept the null hypothesis which states that the $\hat{a} = 0$.

This might look surprising for some of you. But one thing we need to understand is the fact that the sample size (*n*) for this example is very small. Given that small *n*, the ability to reject the false null hypothesis is not so good and that's the reason we are accepting this null hypothesis.

4.2.7 Accuracy of Prediction

The present example has not turned out to be significant. But we will continue to discuss the issues in regression. How accurate we are in predicting Y from X is one of the important issues. We will look at various measures that tell us about the accuracy of prediction. We will continue to use this example considering it as significant even when it is not.

Standard Error of Measurement:

The standard error of estimate provides us an estimate of the error in the estimation. It can be calculated as follows:

$$s_{Y.X}=\sqrt{\frac{\sum(Y-\hat{Y})^2}{n-2}}=\sqrt{\frac{SS_{Residual}}{df}} \qquad \text{(eq. 4.26)}$$

The standard error in our example can be computed using the formula as follows:

$$s_{Y.X} = \sqrt{\frac{SS_{Residual}}{df}} = \sqrt{\frac{60.32}{8}} = 7.54$$

Percentage of Variance Explained: r^2

The r^2 can be used as a measure of amount of variance X explained in Y. This shows the proportion of variance explained from total variance. Look at the following equation.

$$r^2 = \frac{SS_{Regression}}{SS_{Total}} \qquad \text{(eq. 4.27)}$$

$$r^2 = \frac{\sum(\hat{Y} - \bar{Y})}{\sum(Y - \bar{Y})} \qquad \text{(eq. 4.28)}$$

$$r^2 = \frac{\sum(\hat{Y} - \bar{Y})}{\sum(Y - \bar{Y})} = \frac{25.68}{86} = 0.299$$

Which means that 29.9 percent variance in Y is explained by X. This 'explained variance' around 30 percent is a good amount of variance considering the unreliability of psychological variables.

Indeed, the square root of the r^2, will give us the correlation between the X and Y.

Proportional Improvement in Prediction

The Proportional Improvement in Prediction (PIP) is one of the measure of accuracy. It is calculated as follows:

$$\text{PIP} = 1 - \sqrt{(1 - r^2)} \qquad \text{(eq. 4.29)}$$

In case of our example,

$$\text{PIP} = 1 - \sqrt{(1 - r^2)} = 1 - \sqrt{(1 - .299} = 0.162$$

The PIP value for our example is 0.162. So the proportional improvement in prediction is .162.

4.2.8 Assumption Underlying Regression

Some of the important assumptions for doing the regression analysis are as follows:

i) Independence among the pairs of score.

   This assumption implies that the scores of any two observations (subjects in case of most of psychological data) are not influenced by each other. Each pair of observation is independent. This is assured when different subjects provides different pairs of observation.

ii) The variance of the error terms is constant for each value of X.

iii) The relationship between X and Y is linear.

iv) The error terms follow the normal distribution with a mean zero and variance one.

v) Independence of Error Terms. The error terms are independent. They are uncorrelated.

vi) The population of X and the population of Y follow normal distribution and the population pair of scores of X and Y has a normal bivariate distribution.

This assumption(vi) states that the population distribution of both the variables (X and Y) is normal. This also means that the pair of scores follows bivariate normal distribution. This assumption can be tested by using statistical tests for normality.

4.2.9 Interpretation of Regression

The linear regression analysis provides us with lot of information about the data. This information need to be carefully interpreted. The intercept (á), the slope (â), the r^2, the *F*-value, need to be interpreted. Let us take these one by one .

The Intercept (á)

The intercept of regression line is the point at which regression line passes through the y-axis. This point is called as *a* in the sample and á in the population.

One straightforward interpretation of the *a* is it is a regression constant. It is that value, which we need to add into *b*X in order to get the predicted value of Y.

The other way of understanding the intercept is, intercept of regression line is that value of Y when the X value is zero. This interpretation looks intuitive. The correctness of this interpretation depends on whether we have sufficient X values near zero. In our example, the X was Stigma scores. The lowest value of the stigma scores was 49. Obviously we do not have any scores of X that are near zero. So the interpretation is unwarranted. This is due to two reasons.

i) We have not taken the complete range of the X values since we are studying the group of patients.

ii) The real zero X value is almost near impossible and the value of intercept 6.144, which is a Y-value when X is zero, is defiantly not possible.

iii) Nobody would miss -6 appointments. The best is not a single appointment is missed. So this interpretation is not applicable.

Slope (â):

The slope parameter is most important part of regression. The slope is called as regression coefficient. This also has straightforward interpretation.

The slope is that change in the Y when X changes by one unit.

So *rate of change* interpretation is common interpretation of the slope.

In our example, the slope value is 0.1991. This means that if the score on Stigma scale increases by one unit, the number of appointments missed will change by a value of .20.

This would also mean that as the score increases by 5 scale points on the Stigma scale, the person is likely to miss one appointment.

The r^2.:

The r^2 is the value that gives us the percentage of the variance X explains in Y.

The value is .299 in our example.

This means that roughly 30 percent variance in Y can be explained by X.

This can also be understood as proportional reduction in error.

Since we obtain r^2 by dividing the $SS_{Total} - SS_{Error}$ by the $SS_{Total.}$

This tells us about how much of the error is reduced.

The F ratio

The *F*-statistics is computed to test a null hypothesis â = 0.

If the *F*-value is statistically significant then the null hypothesis â = 0 is rejected.

Otherwise one has to accept the null hypothesis. If the null hypothesis is accepted, then there is no need to do the rest of the statistics.

This clearly means that X cannot linearly predict the Y.

However, in our example, the sample size is too small.

So the power of the statistical test is also small.

4.3 STANDARDISED REGRESSION ANALYSIS AND STANDARDISED COEFFICIENTS

We have learned about doing the regression analysis with X and Y. In previous chapters we have also learned to calculate the Z-scores. The Z is a standard score of a variable. To remind you, the Z can be calculated for each of the variable with the formula given below:

$$Z = \frac{X - \bar{X}}{S}$$

The mean of the Z is zero and the standard deviation is one.

Now, instead of predicting Y from X, we calculate the Z scores for both X and Y.

They will be denoted as Z_X and Z_Y.

Now we carry out the regression on standard variables than on unstandardised variables.

The regression equation will be

$$Z_Y = a + bZ_X + e \qquad \text{(eq. 4.30)}$$

Now, the intercept term is completely redundant in this equation because when we take the standard variable (that is Z) then the Y-intercept of the regression line is by default becomes *zero*. so the equation reduces to

$$Z_Y = bZ_X + e \qquad \text{(eq. 4.31)}$$

The beta value obtained in this regression equation is quite interesting.

Let us recall the correlation coefficient.

The correlation coefficient $r = Cov_{XY} / S_X S_Y$.

Now, with both the variable being standardised, the S_X, S_Y and S_X^2, will all be equal to one.

The slope for regression is calculated as $b = Cov_{XY} / S_X^2$.

Now you will realise that the slope (b) is equal to the r, correlation coefficient.

4.4 MULTIPLE REGRESSION

When we have multiple predictors than a single predictor variable, the regression carried out is called as multiple regression.

So we have a dependent variable and a set of independent variables. Suppose we have X_1, X_2, X_3, … up to X_k as k independent variables, and Y as a dependent variable, then the regression equation for sample can be written as:

$$Y = a + b_1 X_1 + b_2 X_2 + ... + b_k K_k \qquad \text{(eq. 4.32)}$$

The same equation for the population can be written as

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k K_k \qquad \text{(eq. 4.33)}$$

Look at the following data. The data is about three variables, number of appointments missed, stigma scores, and the distance between the hospital and home.

Generally, one would expect that if the stigma is high, then the appointments would be missed. Similarly if the hospital is far away, then the appointments may be missed.

Table 4.7: Table of the data for appointments missed, stigma scores and distance of the hospital from home for 10 patients

Appointments Missed (Y)	Stigma Scores (X_1)	Distance of the hospital (X_2)
2	40	2
3	43	5
4	45	4
5	46	7
6	60	9
7	63	5
8	69	2
9	54	8
11	70	6
11	62	9

The equation for which we carry out the regression analysis is as follows:

$$\text{Appointmnets Missed } (Y) = a + b_1 \text{ Stigma Score} + b_2 \text{ Distance from Home} + e$$

We will solve the numerical for this problem. I shall directly provide you with the answer.

The Multiple R^2 for this problem is 0.81. which means that 81 percent information in appointments missed is explained by these two variables.

The adjusted value for the same is .76.

The value of intercept is –7.88.

The slope for stigma is 0.22 and

The slope for distance is 0.40.

The results of significance testing are as follows:

Table 4.8: Table showing the significance testing and the ANOVA summary

Source	Sum of Squares	*Df*	Mean Square	F	Sig.
Regression	73.279	2	36.640	14.981	.003
Residual	17.121	7	2.446		
Total	90.400	9			

The obtained F-value tells us that the overall model we have tested for is turning out to be significant. We can actually test the significance of each of the b separately. When that is done, the b of stigma turned out to be significant ($t = 4.61$, $p < .01$) but the distance did not ($t = 1.93$, $p > .05$).

Here too the size of the sample appears to be the problem leading to non significant results.

The multiple regression equation can be solved hierarchically or directly.

When the equation is solved directly, all the predictors are entered into the equation simultaneously.

When the equation is solved hierarchically, then the predictors are entered one after another depending on the theory or simply depending on their statistical ability to predict the Y.

The multiple regression is very useful technique in psychological research.

4.5 LET US SUM UP

Now we know how to solve the problem of prediction in psychological research. We can develop suitable regression equation and test it against the data. We can test the predictability, amount of information in dependent explained by independent, etc. this technique is very informative.

When we do regression, it does not mean that the causality appears in the equation. It is not a function of statistics. It has to come from theory.

We now know how to set up a multiple regression equation. Though we do not know how to do the calculations, we can understand the results of multiple regression.

4.6 UNIT END QUESTIONS

Given below are some problems with Answers

1) A researcher was interested in predicting marks obtained in the first year of the college form the marks obtained in the high school. He collected data of 15 individuals which is given below. Find out the Independent Variable and Dependent Variable.

   Write regression equation, calculate *a* and *b*, plot the scatter and straight line, write null and alternative hypothesis, determine significance, and comment on the accuracy of the prediction.

School marks	College marks
67	65
45	50
65	60
60	71
55	54
53	49
59	58
64	69
67	75
69	73
70	64
58	66
63	62
71	65
74	78

2) A researcher was interested in predicting general satisfaction of people form perceived social support. She collected data of 10 individuals which is given below. Find out the IV and DV, Write regression equation, calculate *a* and *b*, plot the scatter and straight line, write null and alternative hypothesis, determine significance, and comment on the accuracy of the prediction.

Satisfaction with Life	Perceived Social Support
7	7
6	6
5	6
8	3
9	6
7	4
6	4
3	2
11	9
8	5

3) A researcher was interested in predicting stage performance form social anxiety. She collected data of 10 individuals which is given below. Find out the IV and DV, Write regression equation, calculate *a* and *b*, plot the scatter and straight line, write null and alternative hypothesis, determine significance, and comment on the accuracy of the prediction.

Stage Performance	Social Anxiety
9	11
7	9
6	11
10	7
10	11
9	9
9	8
5	7
14	13
10	9

4) A researcher was interested in predicting attitude to working condition form affective commitment to job. She collected data of 12 individuals which is given below. Find out the IV and DV, Write regression equation, calculate *a* and *b*, plot the scatter and straight line, write null and alternative hypothesis, determine significance, and comment on the accuracy of the prediction.

Attitude to Work	Affective Commitment
5	10
7	13
4	8
5	9
7	14
9	16
3	10
2	6
8	16
7	13
6	9
9	8

Answers:

1) $r = .78$, $r^2 = .608$, $a = 9.27$, $b = .87$, $SS_{Regression} = 641.75$, $SS_{Residual} = 413.19$, $F = 20.19$.

2) $r = .64$, $r^2 = .41$, $a = 3.41$, $b = .69$, $SS_{Regression} = 17.98$, $SS_{Residual} = 26.02$, $F = 5.53$.

3) $r = .51$, $r^2 = .26$, $a = 2.7$, $b = .65$, $SS_{Regression} = 14.67$, $SS_{Residual} = 42.22$, $F = 2.78$.

4) $r = .67$, $r^2 = .45$, $a = .958$, $b = .458$, $SS_{Regression} = 25.21$, $SS_{Residual} = 30.79$, $F = 8.19$.

4.7 SUGGESTED READINGS

Garrett, H.E. (19). *Statistics In Psychology And Education*. Goyal Publishing House, New Delhi.

Guilford, J.P.(1956). *Fundamental Statistics in Psychology and Education.* McGraw Hill Book company Inc. New York.

UNIT 1 CHARACTERISTICS OF NORMAL DISTRIBUTION

Structure

1.0 Introduction

1.1 Objectives

1.2 Normal Distribution/ Normal Probability Curve

- 1.2.1 Concept of Normal Distribution
- 1.2.2 Concept of Normal Curve
- 1.2.3 Theoretical Base of the Normal Probability Curve
- 1.2.4 Characteristics or Properties of Normal Probability Curve (NPC)

1.3 Interpretation of Normal Curve/ Normal Distribution

1.4 Importance of Normal Distribution

1.5 Applications/ Uses of Normal Distribution Curve

1.6 Table of Areas Under the Normal Probability Curve

1.7 Points to be Kept in Mind Consulting Table of Area Under Normal Probability Curve

1.8 Practical Problems Related to Application of the Normal Probability Curve

1.9 Divergence in Normality (The Non-Normal Distribution)

1.10 Factors Causing Divergence in the Normal Distribution/Normal Curve

1.11 Measuring Divergence in the Normal Distribution/ Normal Curve

- 1.11.1 Measuring Skewness
- 1.11.2 Measuring Kurtosis

1.12 Let Us Sum Up

1.13 Unit End Questions

1.14 Suggested Readings

1.0 INTRODUCTION

So far you have learnt in descriptive statistics, how to organise a distribution of scores and how to describe its shape, central value and variation. You have used histogram and frequency polygon to illustrate the shape of a frequency distribution, measures of central tendency to describe the central value and measures of variability to indicate its variation. All these descriptions have gone a long way in providing information about a set of scores, but we also need procedures to describe individual scores or cutting point scores to categorize the entire group of individuals on the basis of their ability or the nature of test paper, which a psychometerician or teacher has used to assess the outcomes of the individual on a certain ability test. For example, suppose a teacher has administered a test designed to appraise the level of achievement and a student has got some score on the test. What did that score mean? The obtained score has some meaning only with respect to other scores either the teacher may be interested to know how many students lie within the certain range

of scores? Or how many students are above and below certain referenced score? Or how many students may be assign A, B, C, D etc. grades according to their ability?

To have an answer to such problems, the curve of Bell shape, which is known as Normal curve, and the related distribution of scores, through which the bell shaped curve is obtained, generally known as Normal Distribution, is much helpful.

Thus the present unit presents the concept, characteristics and use of Normal Distributions and Normal Curve, by suitable illustrations and explanations.

1.1 OBJECTIVES

After reading this unit, you will be able to:

- Explain the concept of normal distribution and normal probability curve;
- Draw the normal probability curve on the basis of given normal distribution;
- Explain the theoretical basis of the normal probability curve;
- Elucidate the Characteristics of the normal probability curve and normal distribution;
- Analyse the normal curve obtained on the basis of large number of observations;
- Describe the importance of normal distribution curve in mental and educational measurements;
- Explain the applications of normal curve in mental measurement and educational evaluation;
- Read the table of area under normal probability curve;
- Compare the Non-Normal with normal Distribution and express the causes of divergence from normalcy; and
- Explain the significance of skewness and kurtosis in the mental measurement and educational evaluation.

1.2 NORMAL DISTRIBUTION/ NORMAL PROBABILITY CURVE

1.2.1 Concept of Normal Distribution

Carefully look at the following hypothetical frequency distribution, which a teacher has obtained after examining 150 students of class IX on a Mathematics achievement test.

Table 1.2.1: Frequency distribution of the Mathematics achievement test scores

Class Intervals	Tallies	Frequency
85 – 89	I	1
80 – 84	II	2
75 – 79	IIII	4
70 – 74	𝍸 II	7
65 – 69	𝍸 𝍸	10
60 – 64	𝍸 𝍸 𝍸 I	16
55 – 59	𝍸 𝍸 𝍸 𝍸	20
50 – 54	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30
45 – 49	𝍸 𝍸 𝍸 𝍸	20
40 – 44	𝍸 𝍸 𝍸 I	16
35 – 39	𝍸 𝍸	10
30 – 34	𝍸 II	7
25 – 29	IIII	4
20 – 24	II	2
15 – 19	I	1
	Total	**150**

Are you able to find some special trend in the frequencies shown in the column 3 of the above table? Probably yes! The concentration of maximum frequencies (f = 30) lies near a central value of distribution and frequencies gradually tapper off symmetrically on both the sides of this value.

1.2.2 Concept of Normal Curve

Now, suppose if we draw a frequency polygone with the help of above distribution, we will have a curve as shown in the fig. 1.2.1

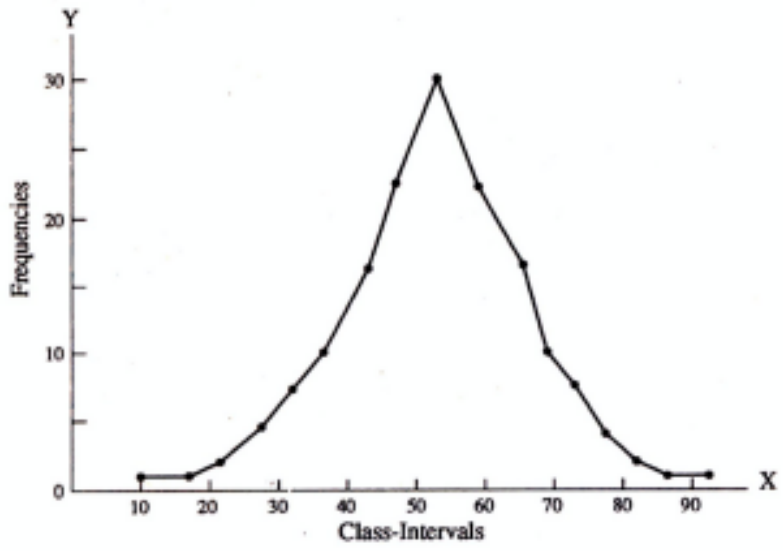


Fig. 1.2.1: Frequency Polygon of the data given in Table 1.2.1

The shape of the curve in Fig. 1.2.1 is just like a 'Bell' and is symmetrical on both the sides.

If you compute the values of Mean, Median and Mode, you will find that these three are approximately the same (M = 52; Md = 52 and Mo = 52)

This Bell shaped curve technically known as Normal Probability Curve or simply Normal Curve and the corresponding frequency distribution of scores, having just the same values of all three measures of central tendency (Mean, Median and Mode) is known as Normal Distribution.

Many variables in the physical (e.g. height, weight, temperature etc.) biological (e.g. age, longevity, blood sugar level and behavioural (e.g. Intelligence; Achievement; Adjustment; Anxiety; Socio-Economic-Status etc.) sciences are normally distributed in the nature. This normal curve has a great significance in mental measurement. Hence to measure such behavioural aspects, the Normal Probability Curve in simple terms Normal Curve worked as reference curve and the unit of measurement is described as σ (Sigma).

1.2.3 Theoretical Base of the Normal Probability Curve

The normal probability curve is based upon the law of Probability (the various games of chance) discovered by French Mathematician Abraham Demoiver (1667-1754). In the eighteenth century, he developed its mathematical equation and graphical representation also.

The law of probability and the normal curve that illust-rates it are based upon the law of chance or the probable occurrence of certain events. When any body of observations conforms to this mathematical form, it can be represented by a bell shaped curve with definite characteristics.

1.2.4 Characteristics or Properties of Normal Probability Curve (NPC)

The characteristics of the normal probability curve are:

1) **The Normal Curve is Symmetrical:** The normal probability curve is symmetrical around it's vertical axis called ordinate. The symmetry about the ordinate at the central point of the curve implies that the size, shape and slope of the curve on one side of the curve is identical to that of the other. In other words the left and right halves to the middle central point are mirror images, as shown in the figure given here.

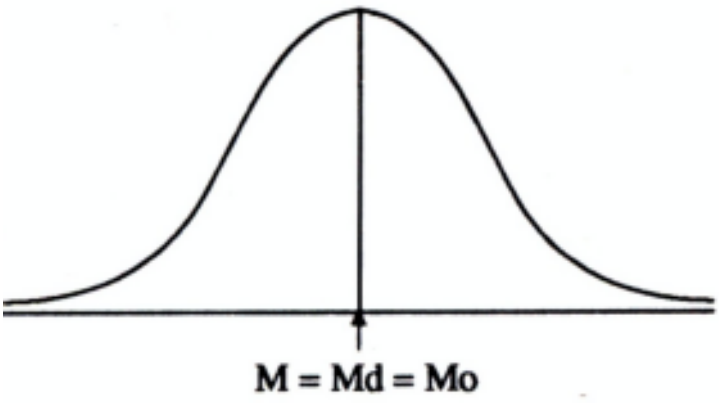


Fig. 1.2.2

2) **The Normal Curve is Unimodel:** Since there is only one maximum point in the curve, thus the normal probability curve is unimodel, i.e. it has only one mode.

3) **The Maximum Ordinate occurs at the Center:** The maximum height of the ordinate always occur at the central point of the curve, that is the mid-point. In the unit normal curve it is equal to 0.3989.

4) **The Normal Curve is Asymptotic to the X Axis:** The normal probability curve approaches the horizontal axis asymptotically; i.e. the curve continues to decrease in height on both ends away from the middle point (the maximum ordinate point); but it never touches the horizontal axis. Therefore its ends extend from minus infinity (- ∞) to plus infinity (+ ∞).

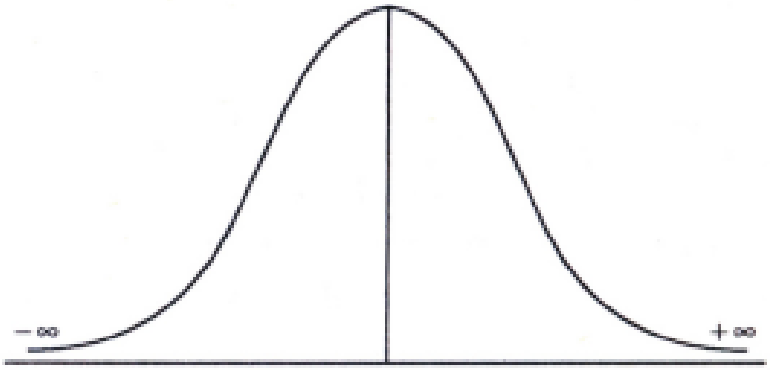


Fig. 1.2.3

5) **The Height of the Curve declines Symmetrically:** In the normal probability curve the height declines symmetrically in either direction from the maximum point.

6) **The Points of Influx occur at point ±1 Standard Deviation (± 1 σ):** The normal curve changes its direction from convex to concave at a point recognised as point of influx. If we draw the perpendiculars from these two points of influx of the curve to the horizontal X axis; touch at a distance one standard deviation unit from above and below the mean (the central point).

7) **The Total Percentage of Area of the Normal Curve within Two Points of Influxation is fixed:** Approximately 68.26% area of the curve lies within the limits of ± 1 standard deviation (± 1 σ) unit from the mean.

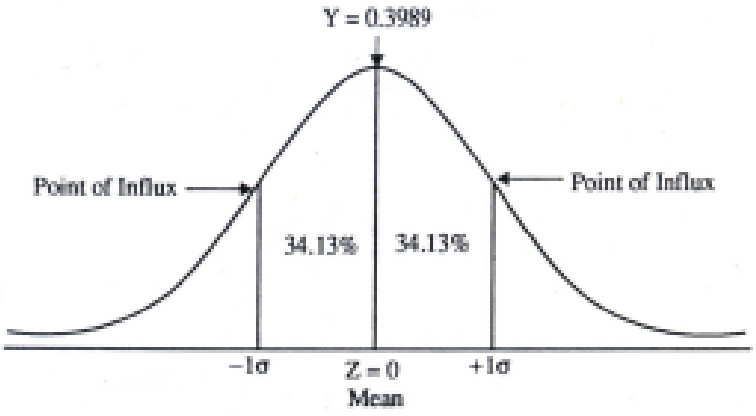


Fig. 1.2.4

8) **The Total Area under Normal Curve may be also considered 100 Percent Probability:** The total area under the normal curve may be considered to approach 100 percent probability; interpreted in terms of standard deviations. The specified area under each unit of standard deviation are shown in this figure.

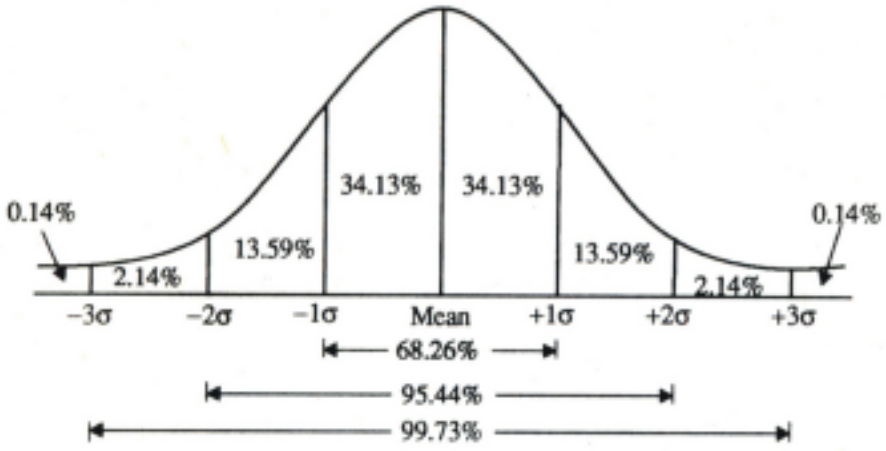


Fig. 1.2.5: The Percentage of the Cases Failing Between Successive Standard Deviation in Normal Distribution

9) **The Normal Curve is Bilateral:** The 50% area of the curve lies to the left side of the maximum central ordinate and 50% of the area lies to the right side. Hence the curve is bilateral.

10) **The Normal Curve is a mathematical model in behavioural Sciences Specially in Mental Measurement:** This curve is used as a measurement scale. The measurement unit of this scale is ± 1σ (the unit standard deviation).

Self Assessment Questions

1) Define a Normal Probability Curve.

.....

.....

.....

.....

2) Write the properties of Normal Distribution.

.....

.....

.....

.....

3) Mention the conditions under which the frequency distribution can be approximated to the normal distribution.

.....

.....

.....

.....

4) In a distribution what percentage of frequencies are lie in between

(a) -1 σ to + 1 σ

(b) -2 σ to + 2 σ

(c) -3 σ to + 3 σ

.....

.....

.....

.....

5) Practically, why are the two ends of normal curve considered closed at the points ±3 σ of the base.

.....

.....

.....

.....

1.3 INTERPRETATION OF NORMAL CURVE/ NORMAL DISTRIBUTION

What do normal curve/ normal distribution indicate? Normal curve has great significance in the mental measurement and educational evaluation. It gives important information about the trait being measured.

If the frequency polygon of observations or measurements of certain trait is a normal curve, it is a indication that

1) The measured trait is normally distributed in the universe.

2) Most of the cases i.e. individuals are average in the measured trait and their percentage in the total population is about 68.26%.

3) Approximately 15.87% (50-34.13%) of cases are high in the trait measured.

4) Similarly 15.87% of cases are low in the trait measured.

5) The test which is used to measure the trait is good.

6) The test which is used to measure the trait has good discrimination power as it differentiates between poor, average and high ability group individuals.

7) The items of the test used are fairly distributed in terms of difficulty level.

1.4 IMPORTANCE OF NORMAL DISTRIBUTION

The Normal distribution is by far the most used distribution in inferential statistics because of the following reasons:

1) Number of evidences are accumulated to show that normal distribution provides a good fit or describe the frequencies of occurrence of many variable facts in biological statistics, e.g. sex ratio in births, in a country over a number of years. The anthropometrical data, e.g. height, weight, etc. The social and economic data e.g. rate of births, marriages and deaths. In psychological measurements

e.g. Intelligence, perception span, reaction time, adjustment, anxiety etc. In errors of observation in physics, chemistry, astronomy and other physical sciences.

2) The normal distribution is of great value in educational evaluation and educational research, when we make use of mental measurement. It may be noted that normal distribution is not an actual distribution of scores on any test of ability or academic achievement, but is, instead, a mathematical model. The distributions of test scores approach the theoretical normal distribution as a limit, but the fit is rarely ideal and perfect.

1.5 APPLICATIONS/USES OF NORMAL DISTRIBUTION CURVE

There are number of applications of normal curve in the field of psychology as well as educational measurement and evaluation. These are:

i) To determine the percentage of cases (in a normal distribution) within given limits or scores.

ii) To determine the percentage of cases that are above or below a given score or reference point.

iii) To determine the limits of scores which include a given percentage of cases to determine the percentile rank of an individual or a student in his own group.

v) To find out the percentile value of an individual on the basis of his percentile rank.

vi) Dividing a group into sub-groups according to certain ability and assigning the grades.

vii) To compare the two distributions in terms of overlapping.

viii) To determine the relative difficulty of test items.

1.6 TABLE OF AREAS UNDER THE NORMAL PROBABILITY CURVE

How do we use all the above applications of normal curve in mental as well as in educational measurement and evaluation? It is essential first to know about the Table of areas under the normal curve.

The Table 1.6.1 gives the fractional parts of the total area under the normal curve found between the mean and ordinates erected at various σ (sigma) distances from the mean.

The normal probability curve table is generally limited to the areas under unit normal curve with N = 1, σ = 1. In case, when the values of N and σ are different from these, the measurements or scores should be converted into sigma scores (also referred to as standard scores or z scores). The process is as follows :

$$z = \frac{X - M}{\sigma} \quad \text{or} \quad z = \frac{x}{\sigma}$$

In which: z = Standard Score X = Raw Score

M = Mean of X Scores σ = Standard Deviation of X Scores

The table of areas of normal probability curve are then referred to find out the proportion of area between the mean and the z value.

Though the total area under the N.P.C. is 1, but for convenience, the total area under the curve is taken to be 10,000 because of the greater ease with which fractional parts of the total area, may be then calculated.

The first column of the table, x/σ gives distance in tenths of σ measured off on the base line for the normal curve from the mean as origin. In the row, the x/σ distance are given to the second place of the decimal.

To find the number of cases in the normal distribution between the mean, and the ordinate erected at a distance of $l\sigma$ unit from the mean, we go down the x/σ column until 1.0 is reached and in the next column under .00 we take the entry opposite 1.0, namely 3413. This figure means that 3413 cases in 10,000; or 34.13 percent of the entire area of the curve lies between the mean and $l\sigma$. Similarly, if we have to find the percentage of the distribution between the mean and 1 .56 σ, say, we go down the x/σ column to 1.5, then across horizontally to the column headed by .06, and note the entry 44.06. This is the percentage of the total area that lies between the mean and 1.56σ.

Table 1.6.1: Fractional parts of the total area (taken as 10,000) under the normal probability curve, corresponding to distance on the baseline between the mean and successive points laid off from the mean in units of standard deviation.

x/σ	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2457	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3290	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3889	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4383	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4988	4984	4984	4985	4985	4986	4986
3.0	4986.5	4986.9	4987.4	4987.8	4988.2	4988.6	4988.9	4989.3	4989.7	4990.0
3.1	4990.3	4990.6	4991.0	4991.3	4991.6	4991.8	4992.1	4992.4	4992.6	4992.9
3.2	4993.129									

3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

Example: Between the mean and a point 1.38 σ $\left(\frac{x}{\sigma} = 1.38\right)$ are found 41.62% of the entire area under the curve.

We have so far considered only σ distances measured in the positive direction from the mean. For this we have taken into account only the right half of the normal curve. Since the curve is symmetrical about the mean, the entries in Table apply to distances measured in the negative direction (to the left) as well as to those measured in the positive direction. If we have to find the percentage of the distribution between mean and -1.28σ, for instance, we take entry 3997 in the column .08, opposite 1.2 in the x/σ column. This entry means that 39.97 percent of the cases in the normal distribution fall between the mean and -1.28σ.

For practical purposes we take the curve to end at points -3σ and +3σ distant from the mean as the normal curve does not actually meet the base line. Table of area under normal probability curve shows that 4986.5 cases lie between mean and ordinate at +3σ. Thus 99.73 percent of the entire distribution, would lie within the limits -3σ and +3σ. The rest 0.27 percent of the distribution beyond ±3σ is considered too small or negligible except where N is very large.

1.7 POINTS TO BE KEPT IN MIND WHILE CONSULTING TABLE OF AREA UNDER NORMAL PROBABILITY CURVE

The following points are to be kept in mind to avoid errors, while consulting the N.P.C. Table.

1) Every given score or observation must be converted into standard measure i.e. Z score, by using the following formula:

$$z = \frac{X - M}{\sigma}$$

2) The mean of the curve is always the reference point, and all the values of areas are given in terms of distances from mean which is zero.

3) The area in terms of proportion can be converted into percentage, and

4) While consulting the table, absolute values of z should be taken. However, a negative value of z shows that the scores and the area lie below the mean and this fact should be kept in mind while doing further calculation on the area. A positive value of z shows that the score lies above the mean *i.e.* right side.

Self Assessment Questions

i) What formula is to use to convert raw score X into standard score i.e. z score.

.....

.....

ii) What is the reference point on the normal probability curve.

iii) Mean value of the z scores is _____

iv) The value of standard deviation of z scores is _____

v) The total area under the N.P.C. is always _____

vi) The negative value of z scores shows that _____

vii) The positive value of z scores shows that _____

1.8 PRACTICAL PROBLEMS RELATED TO APPLICATION OF THE NORMAL PROBABILITY CURVE

Under the caption 1.5 you have studied the Application of normal Distribution/ Normal Curve in mental and educational measurements. Now how the practical problems related to these application are solved you go through the following examples carefully and thoughroly.

1) **To determine the percentage of cases in a normal distribution within given limits of scores.**

Often a psychometrician or psychology teacher is interested to know the number of cases or individuals that lie in between two points or two limits. For example, a teacher may be interested as to how many students of his class got marks in between 60% and 70% in the annual examination, or he may be interested in how many students of his got marks above 80%.

Example 1

An adjustment test was administered on a sample of 500 students of class VIII. The mean of the adjustment scores of the total sample obtained was 40 and standard deviation obtained was 8, what percentage of cases lie between the score 36 and 48, if the distribution of adjustment scores is normal in the universe.

Solution:

In the problem it is given that

$N = 500$

$M = 40$

$\sigma = 8$

We have to find out the total % of the students who obtained score in between 36 and 48 on the adjustment test.

To find the required percentage of cases, first we have to find out the z scores for the raw scores (X) 36 and 48, by using the formula.

$$z = \frac{X - M}{\sigma}$$

$\therefore$ z score for raw score 36 is

$$z_1 = \frac{36-40}{8} =$$

or $z_1 = -0.5\ \sigma$

Similarly z score for raw score 40 is

$$z_2 = \frac{48-40}{8} =$$

or $z_2 = +1\ \sigma$

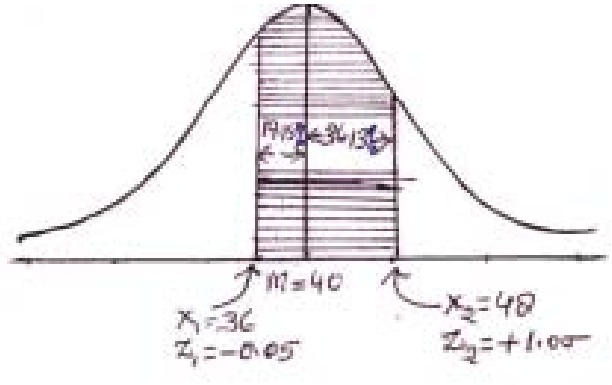


Fig. 1.8.1

According to table of area under Normal Probability curve (N.P.C.) *i.e.* Table No. 1.6.1 the total area of the curve lie in between M to +1σ is 34.13 and in between M to -0.56 is 19.15.

$\therefore$ The total area of the curve in between -0.5 σ to +1 σ is 19.15 + 34.13 = 53.28

Thus the total percentage of students who got scores in between 36 and 48 on the adjustment test is 53.28 (Ans.)

Example 2

A reading ability test was administered on the sample of 200 cases studying in IX class. The mean and standard deviation of the reading ability test score was obtained 60 and 10 respectively. Find how many cases lie in between the scores 40 and 70. Assume that reading ability scores are normally distributed.

Solution:

Given N = 200

M = 60

σ = 10

X_1 = 40 and

X_2 = 70

To find out: The total no. of cases in between the two scores 40 and 70.

To find the required no. of cases, first we have to find out the total percentage of cases lie in between Mean and 40 and mean and 70. See the Fig. 1.8.2 For the purpose, first the given raw scores (40 & 70) should be converted into z scores by using the formula

$$z = \frac{X-M}{\sigma}$$

$$\therefore \quad z_1 = \frac{40-60}{10} =$$

or $z_1 = -2\sigma$

Similarly $z_2 = \frac{70-60}{10} =$

or $z_2 = +1\sigma$

According to Table 1.6.1 the area of the curve in between M and -2σ is 47.72% and in between M and +1σ is 34.13%.

∴ The total area of the curve in between -2σ to +1σ is = 47.72 + 34.13 = 81.85%

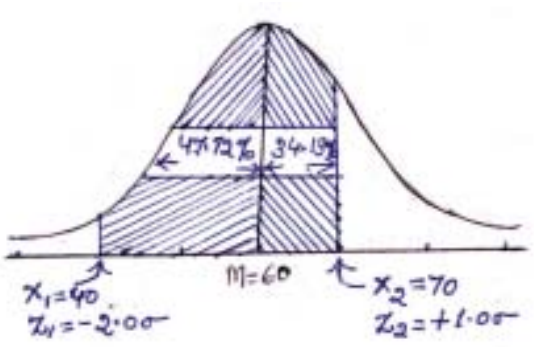


Fig. 1.8.2

Therefore, the total no. of cases in between the two scores 40 and 70 are =

$\frac{81.85\times 200}{100}$ = 163.7 or 164

Thus total no. of cases who got scores in between 40 and 70 are = 164. (Ans.)

2) **To determine the percentage of cases lie above or below a given score or reference point.**

Example 3

An intelligence test was administered on a group of 500 cases of class V. The mean I.Q. of the students was found 100 and the S.D. of the I.Q. scores was 16. Find how many students of class V having the I.Q. below 80 and above 120.

Solution:

Given M = 100, σ = 16, X_1 = 80 and X_2 = 120

To find out : (i) The total no. of cases below 80

(ii) The total no. of cases above 120

To find the required no. of cases first we have to find z scores of the raw scores X_1 = 80 and X_2 = 120 by using the formula

$z = \frac{X-M}{\sigma}$

$z_1 = \frac{80-100}{16} = -\frac{20}{16}$

or z_1 = - 1.25 σ

Similarly,

$z_2 = \frac{120-100}{16} = +\frac{20}{16}$

or z_2 = + 1.25 σ

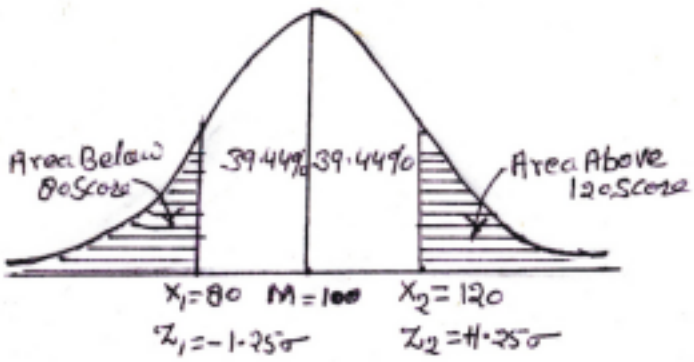


Fig. 1.8.3

According to NPC table (Table 1.6.1) the total percentage of area of the curve lie in between Mean to 1.25 σ is = 39.44.

According to the properties of N.P.C. the 50% area lies below to the mean *i.e.* in left side and 50% area lie above to the mean i.e. in right side.

Thus the total area of NPC curve below M = (100) is = 50 – 39.44 = 10.56

Similarly the total area of NPC curve above M = (100) is = 50 – 39.44 = 10.56

Therefore total cases below to the I.Q. 80 = = 52.8 = 53 Appox.

Similarly Total cases above to the I.Q. 120 = = 52.8 = 53 Appox.

Thus in the group of 500 students of V class there are total 53 students having I.Q. below 80. Similarly there are 53 students who have I.Q. above 120. (Ans.)

3) **To determine the limits of scores which includes a given percentage of cases**

Some time a psychometerician or a teacher is interested to know the limits of the scores in which a specified group of individuals lies. To understand, read the following example-4 and its solution.

Example 4

An achievement test of mathematics was administered on a group of 75 students of class VIII. The value of mean and standard deviation was found 50 and 10 respectively. Find limits of the scores in which middle 60% students lies.

Solution:

Given that, N = 75, M = 50, σ = 10

To find out: Value of the limits of middle 60% cases i.e. X_1 and X_2

As per given condition (middle 60% cases), 30%-30% cases lie left and right to the mean value of the group (see the fig. 1.8.4.)

According to the formula

$$z = \frac{X - M}{\sigma}$$

If the value of M, σ and z is known, the value of X can be find out. In the given problem the value of M and σ are given. We can find out the value of z with the help of the NPC Table No. 1.6.1 as the area of the curve situated right and left to the mean (30%-30% respectively) is also given.

According to the table (1.6.1) the value of z_1 and z_2 of the 30% area is ± 0.84σ

Therefore by using formula

$$z_1 = \frac{X_1 - M}{\sigma}$$

$$-0.84 = \frac{X_1 - 50}{10}$$

or $X_1 = 50 - 0.84 \times 10$

$= 41.60$ or 42

Similarly,

$$z_2 = \frac{X_2 - M}{\sigma}$$

Fig. 1.8.4

$$-0.84 = \frac{X_2 - 50}{10}$$

or $X_2 = 50 + 0.84 \times 10$

$= 58.4$ or 58

Thus $X_1 = 42$

$X_2 = 58$

Therefore, the middle 60% cases of the entire group (75, students) got marks on achievement test of mathematics in between 42 – 58. Ans.

Self Assessment Questions

The observation given in the example 4, i.e. M = 50 and S.D. (σ) = 10

1) Find the limits of the scores middle 30% cases

.....

.....

2) Find the limits of the scores middle 75% cases

.....

.....

3) Find the limits of the scores middle 50% cases

.....

.....

4) **To determine the percentage rank of the individual in his group.**

The percentile rank is defined as the percentage of cases lie below to a certain score (X) or a point.

Some time a psychologist or a teacher is interested to know the position of an individual or a student in his own group on the bases of the trait is measured (for more clarification go through the following example carefully)

Example 5

In a group of 60 students of class X, Sumit got 75% marks in board examination. If the mean of whole class marks is 50 and S.D. is 10. Find the percentile rank of the Sumit in the class.

Solution:

See the fig. 1.8.5. and pay the attention to the definition of percentile given above carefully.

It is clear from the fig. that we have to find out the total percentage of cases (i.e. the area of N.P.C.) lie below to the point X = 75 (See Fig. 1.8.5.)

To find the total required area (shaded part) of the curve, it is essential first to know the area of the curve lie in between the points 50 and 75.

This area can be determined very easily, by taking up the help of N.P.C. Table, i.e. Table No. 1.6.1., if we know the value of z of score 75.

According to the formula

$$z = \frac{X - M}{\sigma}$$

$$z = \frac{75 - 50}{10} = \frac{25}{10}$$

or $= + 2.50\ \sigma$

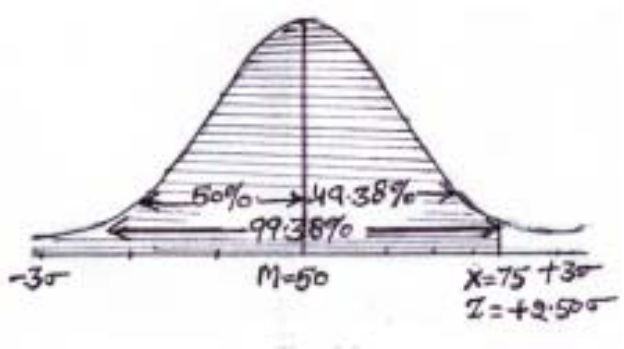


Fig. 1.8.5

According to NPC Table (Table No. 1.6.1) the area of the curve lies M and +2.50 σ is 49.387.

In the present problem we have determined 49.38% area lies right to the mean and 50% area lies to the left of the Mean. (According to the properties of NPC see caption 1.2.4 property no. 9)

Thus according to the definition of percentile the total area of the curve lies below to the point X = 75 is

= 50 + 49.38%

= 99.38% or 99% Approx.

Therefore the percentile rank of the Sumit in the class is 99th. In other words Sumit is the topper student in the class, remaining 99% students lie below to him. (Ans.)

Self Assessment Questions

In a test of 200 items, each correct item has 1 mark.

If M = 100, σ = 10

1) Find the position of Rohit in the group who secured 85 marks on the test.

.....

.....

2) Find the percentile rank of Sunita she got 130 marks on the test.

.....

.....

5) **To find out the percentile value of an individual's percentile rank.**

Some time we are interested to know that the person or an individual having a specific percentile rank in the group, than what is the percentage of score he got on the test paper. To understand, go through the following example and its solution –

Example 6

An intelligence test was administered on a large group of student of class VIII. The mean and standard deviation of the scores was obtained 65 and 15 respectively. On the basis intelligence test if the Ramesh's percentile rank in the class is 80, find what is the score of the Ramesh, he got on the test?

Solution:

Given : M = 65, σ = 15, and PR = 80

To find out : The value of P_{80}

Look at the Fig. No. 1.8.6., as per definition of percentile rank, the 30% area of the curve lie from mean to the point P_{80} and 50% are lie to the left side of the mean.

The z value of the 30% area of the curve lie in between M and P_{80} is = +0.85 σ

(Table No. 1.16)

We know that $z = \frac{X - M}{\sigma}$

or $+ 0.85 = \frac{X - 65}{15}$

or $X = 65 + 15 \times 0.85$

$= 65 + 12.75$

$= 77.75$ or 78 Approx.

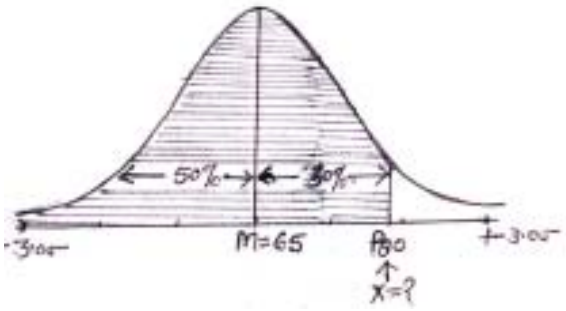


Fig. 1.8.6

Thus Ramesh's intelligence score on the test is = 78 (Ans.)

Self Assessment Questions

1) If M = 100, σ = 10

Find the values of

i) P_{75} = _____

ii) P_{10} = _____

iii) P_{50} = _____

iv) P_{80} = _____

6) **Dividing a group of individuals into sub-group according to the level of ability or a certain trait. If the trait or ability is normally distributed in the universe.**

Some time we are making qualitative evaluation of the person or an individual on the basis of trait or ability, and assign them grades like A, B, C, D, E etc. or 1st grade 2nd grade, 3rd grade etc. or High, Average or Low. For example a company evaluate their salesman as A grade, B grade and C grade salesman. A teacher provides A, B, C etc. grades to his students on the basis of their performance in the examination. A psychologist may classify a group of person on the basis of their adjustment as highly adjusted, Average and poorly adjusted. In such conditions, always there is a question that how many persons or individuals, we have to provide A, B, C, D and E etc. grades to the individuals and categorize them in different groups.

For further clarification go through the following examples:

Example 7

A company wants to classify the group of salesman into four categories as Excellent, Good, Average and Poor on the basis of the sale of a product of the company, to provide incentive to them. If the number of salesman in the company is 100, their average sale of the product per week is 10,00,000 Rs. and standard deviation is Rs. 500/-. Find the number of salesman to place as Excellent, Good, Average and Poor.

Solution:

As per property of the N.P.C. we know that total area of the curve is 6σ over a range of -3σ to $+3\sigma$.

According to the problem, the total area of the curve is divided into four categories.

Therefore area of each category is $6\sigma/4 = \pm 1.5\sigma$. It means the distance of each category from the mean on the curve is 1.5σ respectively.

The distance of each category is shown in the figure 1.8.7

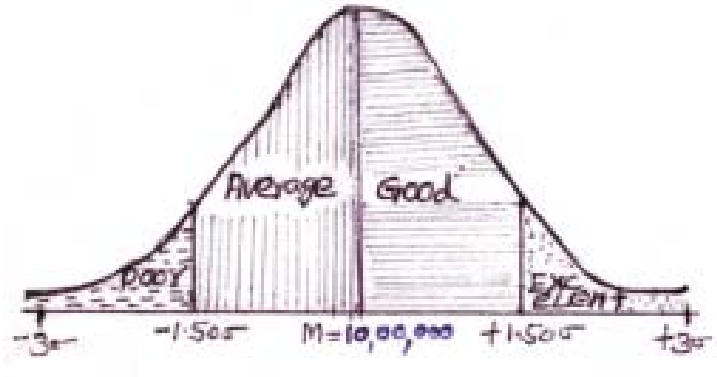


Fig. 1.8.7

i) Total % of salesman in "Good" category

According to N.P.C. Table (Table No. 1.6.1), the

Total area of the curve lies in between M an $+1.5\sigma$ is = 43.32%

$\therefore$ The total % of salesman in "Good" category is 43.32%

ii) Total % of salesman in "Average" category

Total area of the curve lies in between Mean and – 1.5σ is also = 43.32%

∴ The total % of salesman in Average category is = 43.32

iii) Total % of salesman in "Excellent" category

The total area of the curve from M to +3σ and above is

= 50% ..... (As per properties of Normal Curve)

∴ The total % of salesman in the category Excellent is = 50 – 43.32 = 6.68%

iv) Total % of salesman in "Poor" category

The total area of the curve from M to -3σ and below is

= 50% ..... (As per properties of Normal Curve)

∴ The total % of the salesman in the poor category is = 50 – 43.32 = 6.68%

Thus,

i) The number of salesman should place in "Excellent" category

$$= \frac{6.68 \times 100}{100} = 6.68 \text{ or } 7$$

ii) The number of salesman should place in "Good" category

$$= \frac{43.32 \times 100}{100} = 43.32 \text{ or } 43$$

iii) The number of salesman should place in "Average" category

$$= \frac{43.32 \times 100}{100} = 43.32 \text{ or } 43$$

iv) The number of salesman should place in "Poor" category

$$= \frac{6.68 \times 100}{100} = 6.68 \text{ or } 7$$

Total = 100 (Ans.)

Self Assessment Questions

In the above example no. 7 if the salesman are categorised into six categories as excellent, v. good, good average, poor and v. poor. Find the number of salesman in each category as per their sales ability.

.....

.....

Example 8

A group of 1000 applicant's who wishes to take admission in a psychology course. The selection committee decided to classify the entire group into five sub-categories A, B, C, D and E according to their academic ability of last qualifying examination.

If the range of ability being equal in each sub category, calculate the number of applicants that can be placed in groups ABCD and E.

Solution:

Given: N = 1000

To find out: The 1000 cases to be categorised into five categories A, B, C, D, and E.

We know that the base line of a normal distribution curve is considered extend from -3σ to +3σ that is range of 6σ.

Dividing this range by 5 (the five subgroups) to obtain σ distance of each category, i.e. the z value of the cutting point of each category (see the fig. given below)

$$\therefore\ z = \frac{6\sigma}{5} = \pm\ 1.20\ \sigma$$

(It is to be noted here that the entire group of 1000 cases is divided into five categories. The number of subgroups is odd number. In such condition the middle group or middle category (c) will lie equally to the centre i.e. M of the distribution of scores. In other words the number of cases of "c" category or middle category remain half to the left area of the curve from the point of mean and half of the right area of the curve from the mean.

$$\therefore\quad \text{the limits of "c" category is} = \frac{1.2\sigma}{2} = \pm\ 0.60\ \sigma$$

i.e. the "c" category will remain on NPC curve in between the two limits -0.6 σ to +0.6 σ

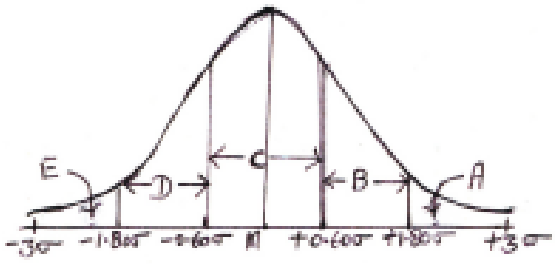


Fig. 1.8.8

Now,

The limits of B category

Lower limit = +0.6 σ

and Upper limit = 0.60 σ + 1.20 σ

or = +1.80 σ

The limits of A category

Lower limit = + 1.8 σ

and Upper limit = + 3 σ and above

Similarly, the limits of D category

Upper limit = - 0.6 σ

Lower limit = (- 0.60 σ) + (-1.20 σ)

or = -1.80 σ

The limits of E category

Upper limit = -1.8 σ

Lower limit = -3σ and below

(For limits of each category see the fig. 1.8.8 carefully)

i) The total % area of the NPC for A category

According to NPC Table (1.6.1) the total % of area in between

Mean to +1.80 σ is = 46.41

∴ The total % of area of the NPC for A category is = 50 – 46.41 = 3.59

ii) The total % Area of the NPC for B category –

According to NPC Table (1.6.1) the total % of Area in between

Mean and + 0.60 σ is = 22.57

∴ The total % area of NPC for B category is = 46.41 – 22.57 = 23.84

iii) The total % area of the NPC for C category –

According to NPC table the total % area of NPC in between

M and + 0.06 σ is = 22.57

Similarly the total % area of NPC in between

M and – 0.06 σ is also = 22.57

∴ The total % area of NPC for C category is = 22.57 + 22.57 = 45.14

iv) In similar way the total % area of NPC for D category is = 23.84

v) The total % area of NPC for E category is = 3.59

Thus the total number of applicants (N = 1000) in –

A category is = $\frac{3.59 \times 1000}{100}$ = 35.9 = 36

B category is = $\frac{23.84 \times 1000}{100}$ = 238.4 = 238

C category is = $\frac{45.14 \times 1000}{100}$ = 451.4 = 452

D category is = $\frac{23.84 \times 1000}{100}$ = 238.4 = 238

E category is = $\frac{3.59 \times 1000}{100}$ = 35.9 = 36

Total= 1000 (Ans.)

Self Assessment Questions

1) In the example 8 if the total applicants are categorised into three categories. Find how many applicants will be the categories A, B and C?

.....

.....

7) **To compare the two distributions in terms of overlapping.**

Example 9

A numerical ability test was administered on 300 graduate boys and 200 graduate girls. The boys Mean score is 26 with S.D. (σ) of 4. The girls' mean. Mean score is 28 with a σ 8. Find the total number of boys who exceed the mean of the girls and total number of girls who got score below to the mean of boys.

Solution:

Given: For Boys, N = 300, M = 26 and σ = 6

For Girls, N = 200, M = 28 and σ = 8

To find: 1- Number of boys who exceed the mean of girls

2- Number of girls who scored below to the mean of boys

As per given conditions, first we have to find the number of cases above the point 28

(The mean of the numerical ability scores of girls) by considering M=26 and σ=6

Second, we to find no. of cases below to the point 26 (The mean score of the boys), by considering M = 28 and 5 – 8 (see the fig. 1.8.9 given below carefully)

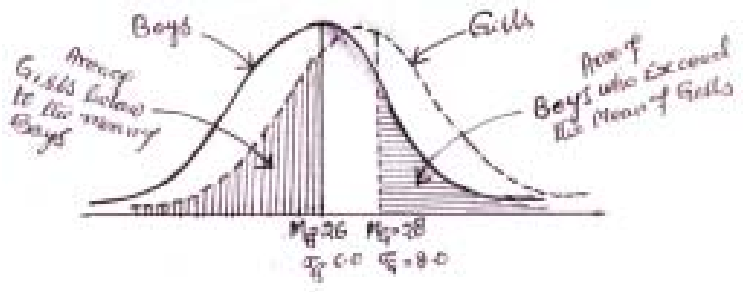


Fig. 1.8.9

1) The z score of X (28) is = $\frac{28-26}{6} = \frac{2}{6}$

or = + 0.33 σ

According to NPC Table (1.6.1) the total % of area of the NPC from M = 26 to + 0.33 σ is = 12.93

$\therefore$ The total % of cases above to the point 28 is = 50 – 12.93 = 37.07

Thus the total number of boys above to the point 28 (mean of the girls) is

$= \frac{37.07 \times 300}{100} = 111.21 = 111$

2) The z score of X = 26 is = $\frac{26-28}{8} = \frac{2}{8} = -0.25\ \sigma$

According to the NPC table the total % of area of the curve in between M = 28 and -0.25 σ is = 9.87

$\therefore$ Total % of cases below to the point 26 is = 50 – 9.87 = 40.13

Thus the total number of girls below to the point 26 (mean of the boys) is

$$= \frac{40.13 \times 200}{100} = 80.26 = 80$$

Therefore,

1) The total number of boys who exceed the mean of the girls in numerical ability is = 111

2) The total number of girls who are below to the mean of the boys is = 88 (Ans.)

Self Assessment Questions

1) In the example given above (Example 9) find.

   i) Number of boys between the two means 26 and 28 _____

   ii) Number of girls between the two means 26 and 28 _____

   iii) Number of boys below to the mean of girls _____

   iv) Number of girls above to the mean of boys _____

   v) Number of boys above to the Md of girls which is 28.20 _____

   vi) Number of girls exceed to the Md of the boys which is 26.20 _____

8) **To determine the relative difficulty of a test items:**

Example 10

In a mathematics achievement test ment for 10th standard class, Q.No. 1, 2 and 3 are solved by the students 60%, 30% and 10% respectively find the relative difficulty level of each Q. Assume that solving capacity of the students is normally distributed in the universe.

Given: The percentage of the students who are solving the test items (Qs) of a question paper correctly.

To Find: The relative difficulty level of each item of the test paper given.

Solution:

First of all we shall mark the relative position of test items on the basis of percentage of students solving the items successfully on the NPC scale.

Q.No.3 of the test paper is correctly solved by the 10% students only. It means 90% students unable to attend the Q.No. 3. On the NPC scale, these 10% cases lies extreme to the right side of the mean (see the fig. given below). Similarly 30% students who are solving Q.No. 2 correctly also lying to the right side of the curve. While the 60% students who are solving Q.No. 1 correctly are lying left side of the N.P.C. curve.

Now, we have to find out the z value of the cut point of the each item (Q.No.) on the NPC base line

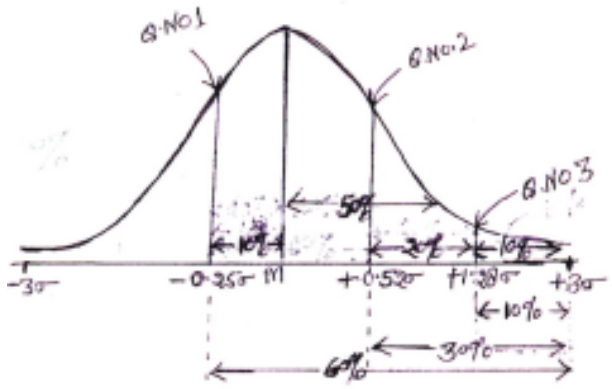


Fig. 1.8.10

i) The z value of the cut point of Q.No. 3

The total percentage of cases lie in between mean and cut point of Q.No. 3 is = (50% - 10%) in right half of NPC

∴ The z value of the right 40% of area of the NPC is = 11.28 σ

ii) The z value of the cut point of Q.No.2

The total percentage of cases lie between the mean and cut point of Q.No. 2 is = 20% (50% - 30%) in right half of NPC

∴ The z value of the right 20% area of the NPC is = + 0.52 σ

iii) The z value of the cut point of Q.No. 3

The total percentage of cases lie between the mean and cut point of Q.No. 3 is = (60% - 50%) in left half of NPC

∴ The z value of the left of 10% of area = - 0.25 σ

Therefore corresponding z value of each item (Q) passed by the students is

Item (Q.No.)	Passed By	z value	Z difference
3	10%	+ 1.28 σ	-
2	30%	+ 0.52 σ	0.76 σ
1	60%	- 0.25 σ	0.77 σ

We may now compare the three questions of the mathematics achievement test, Q.No. 1 has a difficulty value of 0.76 σ higher than the Q.No. 2. Similarly the Q.No. 2 has a difficulty value of 0.77 σ higher than the Q.No. 3. Thus the Q.No. 1, 2 and 3 of the mathematics achievement test are the good items having equal level of difficulty and are quite discriminative. (Ans.)

Self Assessment Question

1) The three test items 1, 2 and 3 of an ability test are solved by 10%, 20% and 30% respectively. What are the relative difficulty values of these items?

.....

.....

.....

1.9 DIVERGENCE IN NORMALITY (THE NON-NORMAL DISTRIBUTION)

In a frequency polygon or histogram of test scores, usually the first thing that strikes one is the symmetry or lack of it in the shape of the curve. In the normal curve model, the mean, the median and the mode all coincide and there is perfect balance between the right and left halves of the curve. Generally two types of divergence occur in the normal curve.

1) Skewness

2) Kurtosis

1) **Skewnes:** A distribution is said to be "skewed" when the mean and median fall at different points in the distribution and the balance *i.e.* the point of center of gravity is shifted to one side or the other to left or right. In a normal distribution the mean equals, the median exactly and the skewness is of course zero ($S_K = 0$).

   There are two types of skewness which appear in the normal curve.

   a) **Negative Skewness :** Distribution said to be skewed negatively or to the left when scores are massed at the high end of the scale, *i.e.* the right side of the curve are spread out more gradually toward the low end *i.e.* the left side of the curve. In negatively skewed distribution the value of median will be higher than that of the value of the mean.

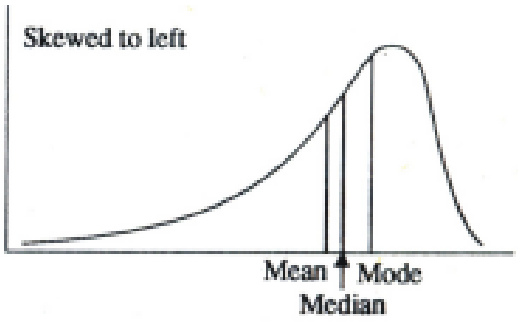


Fig. 1.9.1: Negative Skewness

   b) **Positive Skewness:** Distributions are skewed positively or to the right, when scores are massed at the low; *i.e.* the left end of the scale, and are spread out gradually toward the high or right end as shown in the fig.

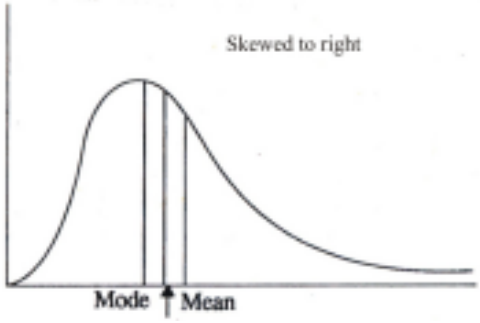


Fig. 1.9.2: Negative Skewness

2) **Kurtosis:** The term kurtosis refers to (the divergence) in the height of the curve, specially in the peakness. There are two types of divergence in the peakness of the curve

   a) **Leptokurtosis:** Suppose you have a normal curve which is made up of a steel wire. If you push both the ends of the wire curve together. What would happen in the shape of the curve? Probably your answer may be that by pressing both the ends of the wire curve, the curve become more peeked *i.e.* its top become more narrow than the normal curve and scatterdness in the scores or area of the curve shrink towards the center.

Thus in a Leptokurtic distribution, the frequency distribution curve is more peaked than to the normal distribution curve.

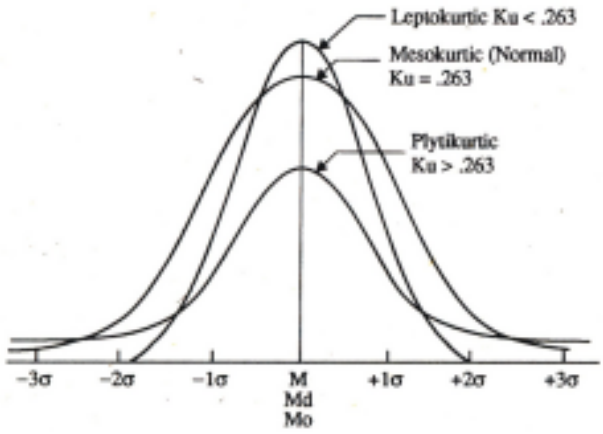


Fig. 1.9.3: Kurtosis in the Normal Curve

   b) **Platykurtosis:** Now suppose we put a heavy pressure on the top of the wire made normal curve. What would be the change in the, shape of the curve? Probably you may say that the top of the curve become more flat than to the normal.

Thus a distribution of flatter Peak than to the normal is known Platykurtosis distribution.

When the distribution and related curve is normal, the vain of kurtosis is 0.263 (KU = 0.263). If the value of the KU is greater than 0.263, the distribution and related curve obtained will be platykurtic. When the value of KU is less than 0.263, the distribution and related curve obtained will be Leptokurtic.

1.10 FACTORS CAUSING DIVERGENCE IN THE NORMAL DISTRIBUTION /NORMAL CURVE

The reasons on why distribution exhibit skewness and kurtosis are numerous and often complex, but a careful analysis of the data will often permit the common causes of asymmetry. Some of common causes are –

1) **Selection of the Sample:** Selection of the subjects (individuals) produce skewness and kurtosis in the distribution. If the sample size is small or sample is biased one, skewness is possible in the distribution of scores obtained on the basis of selected sample or group of individuals.

If the scores made by small and homogeneous group are likely to yield narrow and leptokurtic distribution. Scores from small and highly hetrogeneous groups yield platykurtic distribution.

2) **Unsuitable or Poorly Made Tests:** If the measuring tool or test is unappropriate, or poorly made, the asymmetry is possible in the distribution of scores. If a test is too easy, scores will pile up at the high end of the scale, whereas the test is too hard, scores will pile up at the low end of the scale.

3) **The Trait being Measured is Non-Normal:** Skewness or Kurtosis or both will appear when there is a real lack of normality in the trait being measured, e.g. interest, attitude, suggestibility, deaths in a old age or early childhood due to certain degenerative diseases etc.

4) **Errors in the Construction and Administration of Tests:** The unstandardised with poor item-analysis test may cause asymmetry in the distribution of the scores. Similarly, while administrating the test, the unclear instructions – Error in timings, Errors in the scoring, practice and motivation to complete the test all the these factors may cause skewness in the distribution.

Self Assessment Questions

1) Define the following:

a) Skewness

.....

.....

b) Negative and Positive Skewness

.....

.....

c) Kurtosis

.....

.....

d) Platykurtosis

.....

.....

e) Leptokurtosis

.....

.....

2) In case of normal distribution what should be the value of skewness.

.....

.....

.....

.....

3) In case of normal distribution what should be the value of Kurtosis.

.....

.....

.....

.....

4) What is the significance of the knowledge of skewness and kurtosis to a school teacher?

.....

.....

.....

.....

1.11 MEASURING DIVERGENCE IN THE NORMAL DISTRIBUTION / NORMAL CURVE

In psychology and education the divergence in normal distribution/normal curve have a significant role in construction of the ability and mental tests and to test the representativeness of a sample taken from a large population. Further the divergence in the distribution of scores or measurements obtained of a certain population reflects some important information about the trait of population measured. Thus there is a need to measure the two divergence i.e. skewness and kurtosis of the distribution of the scores.

1.11.1 Measuring Skewness

There are two methods to study the skewness in a distribution.

i) Observation Method

ii) Statistical Method

i) **Observation Method:** There is a simple method of detecting the directions of skewness by the inspection of frequency polygon prepared on the basis of the scores obtained regarding a trait of the population or a sample drawn from a population.

Looking at the tails of the frequency polygon of the distribution obtained, if longer tail of the curve is towards the higher value or upper side or right side to the centre or mean, the skewness is positive. If the longer tail is towards the lower values or lower side or left to the mean, the skewness is negative.

ii) **Statistical Method:** To know the skewness in the distribution we may also use the statistical method. For the purpose we use measures of central tendency, specifically mean and median values and use the following formula.

$$S_K = \frac{3(Mean - Median)}{\sigma}$$

Another measure of skewness based on percentile values is as under

$$S_K = \frac{(P_{90} - P_{10})}{2} - P_{50}$$

Here, it is to be kept in mind that the above two measures are not mathematically equivalent. A normal curve has the value of $S_K = 0$. Deviations from normality can be negative and positively direction leading to negatively skewed and positively skewed distributions respectively.

1.11.2 Measuring Kurtosis

For judging whether a distribution lacks normal symmetry or peakedness; it may detected by inspection of the frequency polygon obtained. If a peak of curve is thin and sides are narrow to the centre, the distribution is leptokurtic and if the peak of the frequency distribution is too flat and sides of the curve are deviating from the centre towards $\pm 4\sigma$ or $\pm 5\sigma$ than the distribution is platikurtic.

Kurtosis can be measured by following formula using percentile values.

$$K_U = \frac{Q}{P_{90} - P_{10}}$$

where Q = quartile deviation i.e.

P_{10} = 10th percentile

P_{90} = 90th percentile

A normal distribution has $K_U = 0.263$. If the value of K_U is less than 0.263 ($K_U < 0.263$), the distribution is leptokurtic and if K_U is greater than 0.263 ($K_U > 0.263$), the distribution is platykurtic.

Self Assessment Questions

1) How we can instantly study the skewness in a distribution.

.....

.....

.....

.....

2) What is the formula to measure skewness in a distribution?

.....

.....

.....

.....

3) What indicates the kurtosis of a distribution?

.....

.....

.....

.....

4) What formula is used to calculate the value of kurtosis in a distribution?

.....

.....

.....

.....

5) How we decide that a distribution is leptokurtic or platykurtic?

.....

.....

.....

.....

1.12 LET US SUM UP

The normal distribution is a very important concept in the behavioural sciences because many variables used in behavioural research are assumed to be normally distributed.

In behavioural science each variable has a specific mean and standard deviation, there is a family of normal distribution rather than just a single distribution. However, if we know the mean and standard deviation for any normal distribution we can transform it into the standard normal distribution. The standard normal distribution is the normal distribution in standard score (z) form with mean equal to 0 and standard deviation equal to 1.

Normal curve is much helpful in psychological and educational measurement and educational evaluation. It provides relative positioning of the individual in a group. It can be also used as a scale of measurement in behavioural sciences.

The normal distribution is a significant tool in the hands of teacher and researcher of psychology and education. Through which he can decide the nature of the distribution of the scores obtained on the basis of measured variable. Also he can decide about his own scoring process which is very lenient or hard; he can Judge the difficulty level of the test items in the question paper and finally he may know about his class, whether it is homogeneous to the ability measured or it is hetrogeneous one.

1.13 UNIT END QUESTIONS

1) Take some frequency distributions and prepare the frequency polygons. Study the normalcy in the distribution. If you will obtained non-normal distribution, determine the type of skewness and kurtosis. Also list down the probable causes associated to the non-normal distribution.

2) Collect the annual examination marks of various subjects of any class and study the nature of distribution of scores of each subject. Also determine the difficulty level of the question papers of each subject.

3) Determine which variables related to cognitive and affective domain of behaviour are normally distributed.

4) As a psychological test constructor or teacher, what precautions are to be considered, while preparing a question paper or test paper.

1.14 SUGGESTED READINGS

Aggarwal, Y.P.: *"Statistical Methods-Concepts, Applications and Computation"*. New Delhi: Sterling Publishers Pvt. Ltd.

Ferguson, G.A.: *"Statistical Analysis in Psychology and Education"*. New York: McGraw Hill Co.

Garrett, H.E. & Woodworth, R.S.: *"Statistics in Psychology and Education"*. Bombay : Vakils, Feffer & Simons Pvt. Ltd.

Guilford, J.P. & Benjamin, F.: *"Fundamental Statistics in Psychology and Education"*. New York: McGraw Hill Co.

Srivastava, A.B.L. & Sharma, K.K.: *"Elementary Statistics in Psychology and Education"*, New Delhi: Sterling Publishers Pvt. Ltd.

UNIT 2 SIGNIFICANCE OF MEAN DIFFERENCES, STANDARD ERROR OF THE MEAN

Structure

2.0 Introduction

2.1 Objectives

2.2 The Concept of Parameters and Statistics and Their Symbolic Representation

- 2.2.1 Estimate
- 2.2.2 Parameter

2.3 Significance and Level of Significance of the Statistics

2.4 Sampling Error and Standard Error

- 2.4.1 Sampling Errors
- 2.4.2 Standard Error

2.5 't' Ratio and 't' Ratio Distribution Table

- 2.5.1 't' Ratio
- 2.5.2 The Sampling Distribution of "t" Distribution

2.6 Standard Error of Sample Statistics – The Sample Mean

- 2.6.1 Meaning of Standard Error of Mean
- 2.6.2 The Standard Error of Mean of Large Sample
- 2.6.3 Degree of Freedom
- 2.6.4 The Standard Error of Means of Small Sample

2.7 Application of the Standard Error of Mean

- 2.7.1 Estimation of the Population of Statistics – The Mpop
- 2.7.2 Determination of the Size of Sample

2.8 Importance and Application of Standard Error of Mean

2.9 The Significance of the Difference Between Two Means

- 2.9.1 Standard Error of the Difference of Two Means and Critical Ratio (CR)
- 2.9.2 Levels of Significance
- 2.9.3 The Null Hypothesis
- 2.9.4 Basic Assumption of Testing of Significance Difference Between the Two Sample Means
- 2.9.5 Two Tailed and One Tailed Test of Significance
- 2.9.6 Uncorrelated (Independent) and Correlated (Dependant) Sample Means

2.10 Significance of the Two Large Independent or Uncorrelated Sample Means

2.11 Significance of the Two Small Independent on Uncorrelated Sample Means

2.12 Significance of the Two Large Correlated Samples

2.13 Significance of Two Small Correlated Means

2.14 Points to be Remember While Testing the Significance in Two Means

2.15 Errors in the Interpretation of the Results, While Testing the Significant Difference Between Two Means

2.16 Let Us Sum Up

2.17 Unit End Questions

2.18 Points for Discussion

2.19 Suggested Readings

2.0 INTRODUCTION

The main function of statistical analysis in behavioural sciences is to draw inferences or made generalisation regarding the population on the basis of results obtained. Therefore the infrential statistics is that branch of statistics which primarily deals with inferences from a sample to a large population from which the sample has been taken. This depends on the fact that how good is the sample estimate. If the sample estimate is not good i.e. having the considerable error or not reliable, we could not be able to draw correct or good inference about the parent population. Thus before to draw inference about the whole population or to made generalisation, it is essential first to determine the reliability or trustworthiness of computed sample mean or other descriptive statistical measures obtained on the basis of a sample taken from a large population.

As an implication of the trustworthiness of the sample measures, we are concerned also with the comparison of two sample estimates with a view to find out if they come from the same population. In other words, the two sample estimates of a given trail of the population do not differ significantly from each other.

Here significant difference means a difference larger than expected by chance or due to sampling fluctuations.

Thus the present unit, highlights the concept of standard error of a sample mean and to compare the two sample means drawn randomly from large population. So that we may be able to test our null hypothesis scientifically, which is made in relation to our experiment or study and to draw the inferences about the population authentically.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- Define and explain the meaning of inference;
- Describe the concept of statistics and parameters;
- Distinguish between statistics and parameters;
- Explain the meaning of significance, significance level;
- Elucidate their role and importance to draw inference and to make generalisation about the population;
- Explain and differentiate between Sampling Error, Measurement Error and Standard;
- Error of Mean value obtained on the basis of a sample from a population;
- Analyse the ‘t’ distribution and its role in inferential statistics;
- Describe the standard error of large and small size sample means;
- Analyse the mean of the population on the basis of the mean of a sample taken from the population with certain level of confidence;
- Determine the appropriate sample size for a experimental study or for a research work Compare the means of two sample means obtained from the same population;
- Differentiate between independent sample means and correlated sample means;
- Test the null hypothesis (Ho) made in relation to an experimental study; and
- Analyse the errors made in relation to testing the null hypothesis.

2.2 THE CONCEPT OF PARAMETERS AND STATISTICS AND THEIR SYMBOLIC REPRESENTATION

Suppose you have administered a verbal test of intelligence on a group of 50 students studying in class VIII of a school of your city. Further, suppose you find the mean I.Q. of this specified group is "105". Can you from this data or information obtained on the relatively small group, say any thing about the I.Q. of all the VIII class students studying in your city. The answer is "Yes" but under certain conditions. The specified condition is "the degree to which sample mean (M) which is also known as "Estimate" represents its parent population mean which is known as "True Mean" or "Parameter". Therefore the two terms Estimates and Parameters are defined as given below.

2.2.1 Estimate

The statistical measurements e.g. measures of central tendency, measures of variations, and measures of relationships obtained on the basis of a sample are known as "Estimates" or Statistics. Symbolically, these are generally represented by using the English alphabets e.g.

Mean = M, Standard Deviation = S.D. or σ, Correlation = r etc.

2.2.2 Parameter

The statistical measurements obtained on the basis of entire population are known as "True Measures" or "Parameters".

Symbolically, these are represented by putting over the bar (-) over corresponding English alphabets or represented by Greek letters e.g.

True Mean or Population Mean = $\overline{M}$ or μ (Mu)

True S.D. or Population S.D. = $\overline{S.D.}$ or $\bar{\sigma}$

True or Population correlation = $\bar{r}$ or η

It is rarely if ever possible to measure all the units or members of a given population. Therefore, practically or for case we draw a small segment of the population with convenient specified number of units or members, which is known as the sample of the population.

Therefore, we do not know the parameters of a given population. But we can under specified condition, forecast the parameters from our sample statistics or estimates with known degree of accuracy.

2.3 SIGNIFICANCE AND LEVEL OF SIGNIFICANCE OF THE STATISTICS

Ordinarily, we draw only a single sample from its parent population. However, our problem becomes one of determining how we can infer or estimate the mean of the population (Mpop) on the basis of the sample mean (M). Thus the degree to which a sample mean (M) represents its parameter is an index of the "Significance" or Trustworthiness of the computed sample mean.

When we draw a sample from the population, the observed statistics or estimate that is the mean of the sample obtained, may be some time large or small to the mean of the population (Mpop). The difference may have arisen "by chance" due to the differences in the composition of our sample, or due to its selection method or the procedure followed in the sample selection. The gap between the two measures sample mean (M) and population mean (Mpop), if is low and negligible the sample mean is considered to be trustworthy and we can forecast or estimate the population mean (Mpop) successfully.

Therefore, a sample mean (M) is statistically trustworthy or significant to forecast the mean of the population (Mpop), depending upon the probability that the difference between the two measures i.e. Mpop and M could have been arisen "by chance".

And the confidence level to which this forecast has been made is known as level of confidence or level of significance.

In simpler terms, the level of significance or level of confidence is a degree to which we accept or reject or predict a happening or incidence with confidence.

There are a number of levels of confidence or levels of significance e.g. 100%, 99%, 95%, 90% ..... 50% etc. In psychology and other behavioural sciences, generally, we consider only two levels of significance viz. the 99% level of significance or level of confidence and 95% level of significance a level of confidence.

The amount 99% and 95% confidence is also termed as 0.01 and 0.05 level of confidence. The 0.01 level means, if we repeatedly draw a sample or conduct an experiment 100 times, only on one occasion, the obtained sample mean or results will fallout side the limits Mpop ± 2.58 S.E.

Here the term S.E. means the standard error exists in the estimate or sample statistics.

Similarly 0.05 level means, if repeatedly draw a sample or conduct an experiment 100 times, only on five occasions the obtained sample mean will fall out side the limits Mpop ± 1.96 S.E.

The value 1.96 and 2.58 have been taken from the 't' distribution or 't' table (P...).

Keeping large size sample in view.

The 0.01 level is more rigorous and higher in terms of standard, as compared to the 0.05 level and would require a high level of accuracy and precision. Hence, if an obtained value (on the basis of a sample or an experiment) is significant at 0.01 level, it is automatically significant at 0.05 level but the reverse is not always true.

2.4 SAMPLING ERROR AND STANDARD ERROR

The score of an individual of a group obtained on a certain test consists of two types of errors (i) measurement error and (ii) statistics error.

In other words,

True Score X_T = observed score or obtained score (X_0) ± error (E)

and error E is = Measurement Error + Statistics Error

The measurement error is caused by a measuring instrument used to measure a trait or variable and personal observation made by the individual on the instrument.

The measurement error is due to the reliability of a test, as no test is perfectly reliable specially in behavioural sciences. In other words no test or a measuring instrument gives us 100% accurate measurement. The personal error is dependent upon the accurate perception and attention of the individual to take observations or measurement on the measuring instrument.

The statistics error refers to the errors of sample statistical measurements or estimates obtained on the basis of a sample drawn from a population.

As it is not possible to have perfectly true representative sample of a population in behaviourable sciences. The statistics error is of two types: (i) Sampling Error (ii) Standard Error of statistics i.e. statistical measurements. Now let us see what are these two errors in detail.

2.4.1 Sampling Errors

Sampling error refers to the difference between the mean of the entire population and the mean obtained of the sample taken from the population.

Thus sampling Error = Mpop ~ M or $\overline{M}$ - M

As the difference is low the mean obtained on the basis of sample is near to the population mean and sample mean is considered to be representing the population mean (or Mpop)

2.4.2 Standard Error

The standard error is nothing but the intra differences in the sample measurements of number of samples taken from a single population.

As the intra differences in the number of sampling observation i.e. the statistics of the same parent population is less and tending to zero, we may say the obtained sample statistics is quite reliable and can be considered as representative of the Mpop or $\overline{M}$.

For more clarification, suppose you wish to determine the I.Q. level of the high school going students studying in the various schools of your district. Is it possible for you to administer the intelligence test on all the high school going students of your district and get the Average I.Q.? Your answer may be certainly not.

The easiest method is to select a sample of 10 schools each from urban and rural areas of your district by using random method and administer the intelligence test to the students studying in high school class of these selected 20 schools in total. In such condition, you may have approximately 20 samples of high school going students and have 20 means of intelligence scores obtained on the intelligence test which you have administered. It is possible that all the mean values you have obtained are not equal. Some may be small and some may be large in their values. Theoretically, there should be no difference in the mean value obtained and all the mean values should be equal as all the samples are taken from the same parent population by using random method of sample selection. Thus the inter variation lies within the values of 20 mean values indicating that the error lies within the various observations taken.

Further, to have the mean I.Q. of all the high school going students you may calculate combined mean or average mean of all 20 means obtained on the basis of samples taken. This obtained combined or average mean value is the I.Q. of the parent population i.e. the high school going students of your district. If you compare all the

20 sample means to the obtained combined mean value i.e. the population mean, you will find that some of sample means are lesser than this Mpop value, and some are higher.

Further one step more, you calculate the difference of these sample means from the population mean obtained, i.e. find (Mpop – M_1), (Mpop – M_2) ..... (Mpop – M_{20}). You will find that some of these difference values are negative and some are positive. The mean of all these differences should be 1.

As the errors are normally distributed in the universe, in simple terms we can say that we have a normal distribution of specific statistics or sample statistical measurement, which is also known as sampling distribution.

Therefore, theoretically the standard error of the statistics (sample statistical measurements) is the standard deviation of the sampling distribution of the statistics and is represented by the symbol $S.E._M$.

Standard Error of sampling measurements on statistics is calculated by using the formula given below :

$$\text{S.E.M or } \sigma_M = \frac{\overline{\sigma}}{\sqrt{N}}$$

Where, S.E.M = Standard Error of sampling measurement

$\overline{\sigma}$ = Standard deviation of the scores obtained from the population.

N = Size of the sample or total number of units in a sample

Look carefully and study the formula given above, you will find, the standard error of any statistics depends mathematically upon two characteristics.

i) the variability or spread of scores around the mean of the population and

ii) the number of units or cases in the sample taken from the population.

As there is low variability in the scores of population, i.e. the population is homogeneous on the trait being measured, and also the number of cases in the sample are too large, the standard error of the statistics is tending to zero.

In the formula, standard error of statistics is directly proportionate to the standard deviation (σ) of the scores of population and inversely proportionate to the size of sample or number of cases in the sample (N).

Thus in brief, it can be said that if the population is homogeneous to the variable or trait being measured and a large size of sample (say more than the 500 units), taken from the population; in such condition the sample drawn will be representative to its parent population and is highly reliable.

Self Assessment Questions

1) Explain the following terms:

   a) Estimate

   .....

   .....

b) Parameter

.....

.....

c) Statistics

.....

.....

d) Sampling Error

.....

.....

e) Measurement Error

.....

.....

f) Standard Error

.....

.....

2) What is the general formula to know the standard errors of the various statistical measures?

.....

.....

.....

3) What do you mean by significance and levels of significance?

.....

.....

.....

4) In behavioural sciences, which levels of confidence are considered

.....

.....

.....

5) What is the difference between significance of statistics and confidence interval for true statistics?

.....

.....

.....

2.5 't' RATIO AND 't' RATIO DISTRIBUTION TABLE

2.5.1 't' Ratio

So far we have studied two types of distributions viz.,

1) Distribution of scores – Normal distribution (unit I)

2) Distribution of 'statistics' or sample statistical measures (p -)

For more clarification, again going back. suppose we have drawn 100 samples of equal size (say n = 500) from a parent population and we calculate the mean value of the scores of a trait of the population obtained from each sample. Thus we have a distribution of 100 means.

Of course all these sample means will not be alike. Some may have comparatively large values and some may have small. If we draw a frequency polygon of the mean values or the "statistics" obtained, the curve will be "Bell – Shaped" i.e. the normal curve and having the same characteristics or properties as the normal probability curve has.

The distribution of statistics values or sample statistical measurements is known as the "sampling – distribution" of the statistics.

The corresponding standard score formula i.e. $z = \frac{X-M}{\sigma}$, may now become as –

$$t = \frac{M-M_{pop}}{S.E._M} \text{ or } \frac{M-\overline{M}}{S.E.\overline{M}}$$

where t = standard score of the sample measures or statistics and termed as "t.Ratio"

M = Mean of specific statistics or sample measure.

$\overline{M}$ or Mpop = Mean of the parameter value of the specific statistics or mean of the specific statistics of the population

$S.E._M$ = Standard Error of the statistics i.e. the standard deviation of the sampling distribution of the statistics.

Actually, t is defined as we have defined the z. It is the ratio of deviation from the mean or other parameter, in a distribution of sample statistics, to the standard error of that distribution.

To distinguish z score of the sampling distribution of sample statistics, we use "t" which is also known as "student's t".

The "t" ratio was discovered by an English statistician, W.S. Gossett in 1908 under the pen name "student". Therefore, the "t" ratio is also know as "student's t" and its distribution is known as "student's t distribution".

As the "t" ratio is the standard score (like z score) with mean = 0 and standard deviation ±1, therefore the t ratio is a deviation of sample mean (M) from population mean ($\overline{M}$ or Mpop).

If this deviation is large the sample statistics mean is not reliable or trustworthy and if the deviation is small, the sample statistics mean is reliable and representative to the mean of its parent population ($\overline{M}$).

2.5.2 The Sampling Distribution of "t" Distribution

Just now we have studied about the sampling distribution of sample statistics and the "t" ratio. Imagine that we have taken number of independent samples with equal size from a population. Let us say we have computed the "t" ratio for every sample statistics with N constant. Thus a frequency distribution of these ratios would be a sampling distribution of "t" and is known as "t" distribution. The mean of all "t" ratios is zero and standard deviation of all "t" ratios i.e. σ is always $\pm\sigma_t$.

It has been observed that if the sample size varies the sampling distribution of "t" also varies, though it is normal distribution. The sampling distribution of "t" may vary in kurtosis. Student's t distribution becomes increasingly leptokurtic as the size of sample decreases.

As the size of sample is tending to be large, the distribution of "t" approaches to the normal distribution. Thus we have a family of "t" distributions, rather to one and the σt values varies on the x axis.

Fisher has prepared a table of "t" distribution having N, i.e. the size of sample different for different levels of significance. The details of the same are given below:

Table 2.5.1 : Table of "t" for use in determining the significance of statistics

Degrees of Freedom	Probability (P) 0.10	0.05	0.02	0.01
1	$t = 6.34$	$t = 12.71$	$t = 31.82$	$t = 63.66$
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.90	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
11	1.80	2.20	2.72	3.11
12	1.78	2.18	2.68	3.06
13	1.77	2.16	2.65	3.01
14	1.76	2.14	2.62	2.98
15	1.75	2.13	2.60	2.95
16	1.75	2.12	2.58	2.92
17	1.74	2.11	2.57	2.90
18	1.73	2.10	2.55	2.88
19	1.73	2.09	2.54	2.86
20	1.72	2.09	2.53	2.84
21	1.72	2.08	2.52	2.83
22	1.72	2.07	2.51	2.82
23	1.71	2.07	2.50	2.81
24	1.71	2.06	2.49	2.80
25	1.71	2.06	2.48	2.79
26	1.71	2.06	2.48	2.78
27	1.70	2.05	2.47	2.77
28	1.70	2.05	2.47	2.76
29	1.70	2.04	2.46	2.76
30	1.70	2.04	2.46	2.75
35	1.69	2.03	2.44	2.72
40	1.68	2.02	2.42	2.71
45	1.68	2.02	2.41	2.69
50	1.68	2.01	2.40	2.68
60	1.67	2.00	2.39	2.66
70	1.67	2.00	2.38	2.65
80	1.66	1.99	2.38	2.64
90	1.66	1.99	2.37	2.63
100	1.66	1.98	2.36	2.63
125	1.66	1.98	2.36	2.62
150	1.66	1.98	2.35	2.61
200	1.65	1.97	2.35	2.60
300	1.65	1.97	2.34	2.59
400	1.65	1.97	2.34	2.59
500	1.65	1.96	2.33	2.59
1000	1.65	1.96	2.33	2.58
∞	1.65	1.96	2.33	2.58

Let us now take an example. Let us say there are 26 subjects. N = 26.

Example: When N = 26, the corresponding degree of freedom (df) is N-1 i.e. 25.

In column 3 at 0.05 level of significance the t value is 2.06.

It means that five times in 100 trials a divergence of sample mean or statistics obtained may be expected at a 2.05 σ to M, that is to its mean population either to its left or right side.

The "t" distribution table has great significance in inferential statistics testing the null hypothesis framed in relation to various experiments made in psychology and education.

2.6 STANDARD ERROR OF SAMPLE STATISTICS – THE SAMPLE MEAN

The standard error of sample statistics or the statistical measurements of a sample has great importance in inferential statistics. With the help of the standard error statistics we can determine the reliability or trustworthiness of the descriptive statistics e.g. proportion percentage, measures of central tendency (mean, median and mode) measures of variability (standard deviation & quartile deviation), Measures of correlation (r, p and R) etc.

For convenience here we discuss only the significance of means which are detained as under :

2.6.1 Meaning of Standard Error of Mean

The Standard error of mean measures the degree to which the mean is affected by the errors of measurement as well as by the errors of Sampling or Sampling fluctuations from one random sample to the other. In other words how dependable is the mean obtained from a sample to its parameter i.e. population Mean (M pop).

Keeping in mind the Sample Size; there are two situations:

i) Large Sample

ii) Small Sample

2.6.2 The Standard Error of Mean of Large Sample

When we say large sample, the number of items in the sample will be more than 30. That is N>30. In such condition the Standard error of the Mean is determined by using the formula given below:

$S.E._M = \sigma / \sqrt{N}$ when N= >30

Where

$S.E._M$= Standard Error of the Mean of the scores of a large sample

σ = Standard deviation of the scores of a population

N = Size of the sample or number of cases in the sample

This formula is used when the population parameter of standard deviation (σ) is known. But in practice it is not possible to have the value of Ïó. Having the situation that the sample is selected from the population by using random method of sample selection, σ can be replaced by the ó. The value of the standard deviation of the scores of the sample taken. Therefore, in the above formula the σ can be replaced by ó.

$S.E._M = ó / \sqrt{N}$ when N >30

where

$S.E._M$ = Standard Error of the Mean of the scores of a Sample

ó = Standard Deviation of the scores of a sample.

N = Number of units or cases in the Sample.

Example 1: A reasoning test was administered on a sample of 225 boys of age group 15 + years. The mean of the scores obtained on the test is 40 and the standard deviation is 12. Determine how dependable the mean of sample is.

Given : N=225. M=40 and ó = 12

To find : The trustworthyness of the sample mean we know that standard error of the mean, when N>30 is determined by using the formula-

$S.E._M = ó / \sqrt{N}$

$S.E._M = 12 / \sqrt{225}$

$= 12 / 15 = 0.80$

Or $S.E_M = 0.80$

i.e. $= 0.80$

Interpretation of the Result

Keeping in mind the logic of sampling distribution, that is if we draw 100 samples, each sample has 225 units from a large population of boys of age group 15+ years, we will have 100 sample means falling into a normal distribution around the Mpop and $ó_M$ (the standard deviation of sampling distribution of Means i.e. the standard error of Mean)

As per properties of Normal Distribution, in 95% cases the sample means will lie within the range of ±1.96 in to the Mpop (see Z table in unit I). Conversely out of 100, the 99 sample means having equal size, will be within the range of ±2.57 (2.57 ×0.80) of the Mpop.

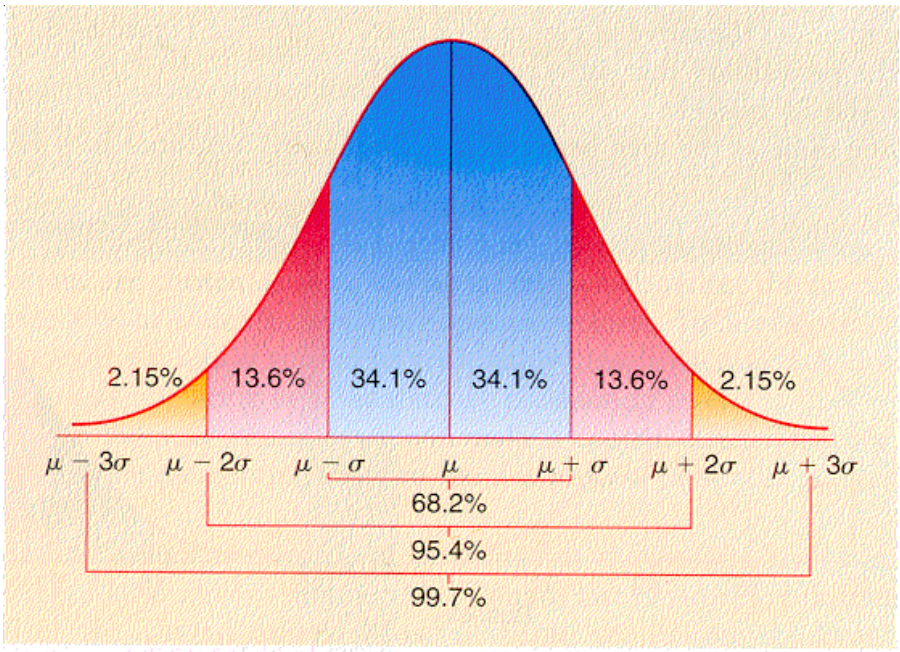


Fig. 2.6.1: Image graph

(**Source:** kmblog.rmutp.ac.th/.../28/normal-distribution/)

From the figure it is quite evident that the S.E. of the M=40. Sample of 225 having

bó = 12 lie within the acceptable region of the N.P.C.(Normal Probability curve). Thus the sample mean obtained is quite trustworthy with the confidence of 95% probability. There are only 5% chances that the sample mean obtained will lie in the area of the rejection of M.P.C.

In simplest term we can say that, there is 95% probability the maximum possibility of the standard error of the sample mean (40) is ± 1.57 (1.96×0.80) which is less than the value of T=1.96 at .05 level of confidence for df=224 (N-1) Thus the obtained sample mean (40) is quite dependable to its Mpop with the confidence level of 95%.

Example 2: In the example 1, suppose in place of N=225, we have a sample of 625 units and the remaining observations are the same. Determine how good an estimate is it of the population mean?

Solution

Given : N=625, M=40 and ó=12

To find : Dependency of sample Mean or reliability of sample mean

We know that

$$ó_M / S.E._M = \frac{\sigma}{\sqrt{N}}$$

$$= \frac{12}{\sqrt{625}}$$

$$= \frac{12}{25} = 0.48$$

Or $S.E._M$=0.48.

Interpretation of Result

The maximum standard error of sample M=40 and $\sum$ =12 having 625 units is ±0.94 (1.96×0.48) at 95% level of confidence which is much less than the value of t.05 = ± 1.96. Therefore, the obtained sample mean is reliable and to the considered as representative to its Mpop at 95% level of confidence.

Self Assessment Questions

1) Compare the two results obtained from Example no 1 and 2 respectively. What you have observed and what is your conclusion.

.....

.....

.....

.....

.....

2) The mean achievement score of a random sample of 400 psychology students is 57 and D.D. is 15? Determine how dependable is the sample mean?

.....

.....

.....

.....

3) A sample of 80 subjects has the mean = 21.40 and standard deviation 4.90. Determine how far the sample mean is trustworthy to its Mpop.

.....

.....

.....

.....

2.6.3 Degree of Freedom

Before to proceed for the standard error of small sample mean, it is imperative here to understand the concept of Degrees of Freedom.

The expression degrees of freedom is abbreviated from the full expression "Degrees of Freedom to Vary". A point in the space has unlimited freedom to move in any direction, but a point on a straight line has only one freedom to move on the line i.e. it is free to move in one dimension only. A point on the plane has two degrees of freedom. A point on a three dimension space having three degree of freedom to move.

This shows that the degree of freedom is associated to the number of restrictions imposed upon the observations.

The degree of freedom is a mathematical concept and is a of key importance in inferential statistics. Almost all test of significance require the calculation of degree of freedom.

When a sample statistics is used to estimate a parameter, the number of degrees of freedom depends upon the number of restrictions placed upon the scores, each restriction reducing one degree of freedom (df).

For example, we have four numbers 4, 5, 8 and 3, the sum of these numbers is 20 which is fixed. e.g. here we have restricted freedom to get sum 20, as we have only to change the values of first three figures and not the last one, as it depends upon the fixed sum 20,

i.e.

$6 + 6 + 9 \quad +\text{———}\text{-}20$

Or $2 + 7 + 10 + \text{———}20$

Or $6 + 7 + 5 \quad +\text{———}20$

In the above sum expressions we have last one freedom to determine the value of

last 4th numeral figure as to get fixed sum 20. Therefore in the above expressions, we are bound to take forcefully the numeral figures 4,1 and 2 respectively.

Like wise, in Statistics, when we calculate Mean or S.D. of a given distribution of scores we lose one degree of freedom to get fixed sums. Therefore we have N-1 df to compute statistical measures , specifically the standard deviation of the scores given and N-2 in case to compute the co-efficient of correlation. It means the degree of freedom is not always (N-1) however, but will vary with the problem and the restrictions imposed.

In the case of very large sample used in behavioural sciences or social sciences no appreciable difference takes place in the value of $ó_M$ by N-1 instead of N. The use of N or N-1 thus remains a matter of arbitrary decision. But in the case of small samples having number of units or cases below 30 $ó_M$ no correction (N-1) has been applied in computation of S.D and thus a considerable variation occurs on the value of $ó_M$ … Therefore, it is imperative to use N-1 in place of N in computation of $ó_M$ of the small sample.

Further you have to study the t" distribution table (table No-2.5.1) very carefully. In the process, you will find that as the size of sample or degree of freedom approaches to 500 or above, the "t" value approaches to the value of 95% and 99% are 1.96 and 2.58 respectively and remain constant. It means the "t" distribution becomes normal distribution or Z-distribution. When the size of sample decreases especially below 30 you will find the "t" values are gradually increasing at 95% level and 99% level considerably. In such condition the $ó_M$ values gives us wrong information and we may interpret the results inappropriately.

2.6.4 The Standard Error of Means of Small Sample

The small sample means, when size of sample (N) is about 30 or less, is treated as small sample. The formula for the standard error of small sample mean score is as follows–

$$S.E._M \text{ on } S_M = ó / \sqrt{N-1}$$

As here $S.E_M$= Standard error of Mean of Small sample

Standard deviation of the population

N = Size of the sample i.e. 30 or below

Note : For practice we replace -ó by ó i.e. standard deviation of the sample. Because of the reason ó is not possible to obtain for whole population.

Example 3: A randomly selected group of 17 students were given a word cancellation test. The mean and S.D obtained for cancelling the words per minute is 58 and 8 respectively. Determine how far sample mean is acceptable to represent the Mean of the population?

Solution

Given : N=17, M=58 and ó = 8

To find : dependency of the sample mean

In the problem the size of sample is less than 30. Therefore to find the standard error of sample mean is

$$S.E_M = \frac{\sigma}{\sqrt{N-1}}$$

$$S.E_M = \frac{8}{\sqrt{17-1}}$$

$$= \frac{8}{\sqrt{16}}$$

$$= 8 / 4 = 2$$

In the "t" table (table no. . 2.5.1) at .01 level, the value of "t" for 16 df is 2.92 and the obtained value of t = 2.00. which is less in comparison to the "t" value given in the able. Therefore, the obtained sample mean (58) is quite trustworthy and representing its Mean population by 99% confidence. There is only one chance out of 100, that sample mean is low or high.

Self Assessment Questions

1) What is the concept of Degree of Freedom (df)?

.....

.....

.....

.....

2) Why we consider the (df) which determining the reliability or trustworthiness of the statistics.

.....

.....

.....

.....

3) What is the difference to calculate the standard error of Mean of Large Size and Small Size samples.

.....

.....

.....

.....

2.7 APPLICATION OF THE STANDARD ERROR OF MEAN

2.7.1 Estimation of the Population Statistics – The Mpop

The wider use or application of $S.E._M$ is to estimate the population Statistical measurements i.e. the M pop. Here we are concerned only with the mean. Therefore we will discuss to estimate the mean of population on the basis of standard error of the mean obtained on either large size sample or small size sample.

Here, it is important to note that the estimation is always in range rather to point estimation because exact single value of any measurement is not possible. For example a student can not forecast with confidence that he will secure 85 to 95 marks out of 100 in statistics in the final examination.

Therefore, estimation of Mpop is always in range rather to point. Thus the limits obtained of Mpop (The lower and upper limits) are also known as Fiduciary limits.

The term Fiduciary limits was used by R.A. Fisher for the confidence interval of parameter and the confidence placed in the interval defined as Fiduciary Probability.

The simplest formula to estimate the Mpop is as under:

Mpop or M = M± 2.58 $ó_M$ ——— (at .01 level of significance)

Mpop or M= M±1.96 $ó_M$ ————(at .05 level of significance)

For more classification study the following examples carefully:-

Example 4: One language test was given to 400 boys of VIII class, the mean of their performance is 56 and the standard deviation is 14. What will be the Mean of the population of 99% level of confidence?

Solution

Given: N=400, M=56 and ó =14

To find out: Estimation of population mean at 99% level of confidence.

We know that Mpop at .01 level or at 99% confidence level is

Mpop .01 or M.01= M±2.58 $ó_M$

Where

M .01: Mean of the population at 99% confidence level of confidence

M : Mean of the sample

$ó_M$: Standard Error of the sample mean.

In the problem the values of Mean and N are known and the value of $ó_M$ is unknown. The value of $ó_M$ can be determined by using the formula.

$$ó_M = \frac{\sigma}{\sqrt{N}}$$

$$\therefore ó_M = \frac{14}{\sqrt{400}} = \frac{14}{\sqrt{20}}$$

Or $ó_M$ = 0.70

Thus $\overline{M}.01 = 56 \pm 2.58 \times 0.70$

$= 56 \pm 1.806$

$= 54.194\text{-}57.806$

Or $\overline{M}.01 = 54\text{-}58$

The Mean of the population at 99% level of confidence will be within the limits 54 to 58. In other words there are 99% chances that the Mean of the population lie within the range 54-58 scores. There is only 1% chance that mean of the population lie beyond this limit.

Example 5: A randomly selected group of 26 VI grade students having a weight of 35 kg and S.D = 10 kg. How well does this value estimate the average weight of all VI grade students at .99 and .95 level of confidence?

Solution

Given : N =26, M = 35kg and ó= 10 kg.

To find out : The fiduciary limits of the population mean at .05 and .01 levels.

In the problem the given sample size is below 30, Therefore to have the standard error of sample mean we will use the formula:

S_M = 2.0

And = N – 1 = 26 – 1

Or df = 25

i) Fiduciary limits of M at .01 level of confidence

By consulting the t table, level of confidence, the value of "t" for 25 df is 2.79

Thus, the Fiduciary limit of M at .01 or 99% level is

$= \bar{M} \pm 2.79\sigma_M$

$= 3.5 \pm 2.79 \times 2.00$

$= 3.5 \pm 5.50$

$\therefore \bar{M}.01 =$ 29.42-38.58 = (9.16)

ii) The fiduciary limits of $\bar{M}$ at .05 level of confidence

$M_{0.05} = M \pm 2.06\ \sigma_M$

$= 35 \pm 2.06 \times 2.0$

$= 35 \pm 4.12$

Or $M_{0.05}$ = 30.88–39.12

i) Thus The Fiduciary Limits of M .01 = 29.42-38.58

ii) The Fiduciary Limits of M.05 = 30.88-39.12

2.7.2 Determination of the Size of Sample

Standard error of the statistics is also used to estimate the sample size for test results. In order to learn how to use the standard error of the statistics you must go through the following examples.

Example 6

If the standard deviation of a certain population (ó) is 20. How many cases would require in a sample in order that standard error of the mean should not miss by 2.

Solution

Given: ó= 20, and S. $E._M$=20.

To Find Out : No of cases in the sample to be selected i.e. to determine the Size of Sample (N)

We know that

$$S.EM = \frac{\sigma}{\sqrt{N}}$$

$$\therefore\ 2 = \frac{20}{\sqrt{N}}$$

$$\text{Or } \sqrt{N} = \frac{20}{2} = 10$$

Or N $= (10)^2$

Or N $= 100$

If the standard error of the sample mean should not be more than 2 in such condition the maximum sample Size Should be 100 i.e. N=100 .

Example 7: The standard deviation of the intelligence scores of an adolescent population is 16. If the maximum acceptable standard error of the mean of the sample should not miss by 1.90, what should be the best sample size at 99% level of confidence?

Solution

Given :ó = 16, SE_M= 1.90

To find out : Sample size which represent its parent population upto the level of 99%.

We know that the Z value of 99% cases is 2.58 (From Z Table)

It means due to chance factors the sample mean would deviate from Mpop by 2.58 $ó_{M.}$ Further in keeping view the measurement and other uncontrolled factors, the measured error in the sample mean we would like to accept is 1.90.

Therefore the maximum error in the sample which we would like to select from the parent population is

$$S.E_M = ó \times \frac{2.58}{\sqrt{N}}$$

$$\text{Or } \sqrt{N} = \frac{\sigma \times 2.58}{SE_M}$$

$$\text{Or N} = \left(\frac{\sigma \times 2.58}{SE_M}\right)^2$$

$$\therefore \text{N} = \left(\frac{16 \times 2.58}{1.90}\right)^2$$

Or N = 472

To have a representative sample up to the level of 99% to the parent population, it is good to have a sample size more than 472 cases.

Self Assessment Questions

1) Given M =26.40, ó = 5.20 and N=100 compute

   The fiduciary limits of True Mean at 99% confidence interval

   The fiduciary limits of Population Mean at .95confidence interval.

   .....

   .....

   .....

   .....

2) The mean of 16 independent observations of a certain magnitude is 100 and S.D is 24. At .05 confidence level what are the fiduciary limits of the True Mean.

   .....

   .....

   .....

   .....

3) Suppose it is known that S.D of the scores in a certain population is 20. How many cases would we in a sample in order that the S.E of the sample mean be σ 2.

   .....

   .....

   .....

   .....

2.8 IMPORTANCE AND APPLICATION OF STANDARD ERROR OF MEAN

The Standard error of statistics has wide use in inferential statistics. It helps the experimenter or researcher in drawing concrete conclusions rather than abstract ones.

The various uses of standard error of the statistics are as under:

Various devices are used for determining the reliability of a sample taken from the large population. The reliability of the sample depends upon the reliability of the statistics, which is very easy to calculate.

The main focus of the standard error of statistics is to estimate the population parameters. No sampling device can ensure that the sample selected from a population may be representative. Thus the formula of the standard error of statistics provides us the limits of the parameters, which may remain in an interval of the prefixed confidence interval.

The method of estimating the population parameters the research work feasible,

where the population is unknown as impossible to measure. It makes the research work economical from the point of crew of time. Energy and money.

Another application of the standard error of the statistics is to determine the size of the sample for experimental study or a survey study.

The last application of the standard error of statistics to determine the significance of difference of two groups to ascertained by eliminating the sampling or change by estimating the sampling or change errors.

2.9 THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO MEANS

Suppose we wish to study the linguistic ability of the two groups say boys and girls of age group 15 years.

First we have to select two large and representative samples from the two different populations. The population of the Boys of age group 15 years old and second, the population of girls of same age as of the Boys has has to be selected.

To administer the linguistic ability test Battery to both the groups of Boys and Girls selected as sample.

To compute the mean values of like scores obtained by the two groups on the linguistic ability test battery and find the difference between them.

After the above procedural steps, suppose that there is a difference in the means of the two groups and the difference is in favour of the girls of age group 15+ years old. Is this evidence sufficient to draw the conclusion that girls are superior in linguistic ability in comparison to the boys having same age level?

Probably the answer to this question may either be "yes" or "No" depending upon further testing of the difference of the means of two groups whether it is statistically significant or not. In other words, it is essential to test further, that how far the difference exists in the mean values of two groups is due to " chance" factor or is it "real and dependable"

This question involves the standard error of the difference that exists between the mean values of the two groups and the same as significant or not. Therefore, in order to test the significance of an obtained difference, we must first have a S.E. of the difference of sample means. Then from the difference between the sample means and the standard error of the difference of sample means (S.E. $_{D.M.}$), we can determine whether a difference probably exists between the population means.

A difference is called *Significant* when the probability is high that it cannot be attributed to chance (i.e. by temporary and accidental factors or by sampling fluctuations) and hence represents a true difference between population means. And a difference is *non-significant* or chance, when it appears reasonably certain that it could be easily have arisen from sampling fluctuations and hence implies no real or true difference between the population means.

Thus the above discussion leads us to conclude that the significance of the difference between two sample means obtained from the two populations either independent or correlated depend upon two factors, viz.,

i) Standard Error of Difference between the two means, and

ii) The levels at which S.E $_{DM}$ is significant.

iii) Standard Error of the difference of two means (Σ_{DM}) and critical ratio (C.R)

2.9.1 Standard Error of the Difference of Two Means and Critical Ratio (CR)

Suppose we have two independent large populations, say A and B, and let us say that we have taken several numbers of samples (say two) from each population. Now if we compute the mean values of the scores of a trait of the two populations, we have 100 sample means obtained from the population A and 100 sample means obtained from the population B, and let us say that we find that there is a difference between the two sample means of population A and B. Thus in this way we have 100 differences of sample means. If we plot the frequency polygon of these hundred samples, certainly we will have a normal curve, and the distribution of the sample mean differences will be known as sampling *Distribution of Mean Differences.*

The standard error of the sample mean differences can be obtained by computing standard deviation of the sampling distribution of mean differences. This can be computed by using the formula:

$$S.E_M \text{ or } \Sigma_{DM} = \frac{\Sigma 1^2 + \Sigma 2^2}{N_1 + N_2} \text{ (In case of two independent population)}$$

Where

Σ_1 = Standard Deviation of the scores of a trait of the sample-1

Σ_2 = Standard Deviation of the scores of a trait of the sample-2

N_1 = Number of cases in sample-1

N_2 = Number of cases in sample-2

After having the standard error of the sample mean differences, the next step is to decide how far the particular sample mean difference is deviating from the two population mean differences ($M_1 \sim M_2$) on the normal probability curve scale. For the purpose we have to calculate Z score of the particular two sample mean differences, using the formula

$$Z = \frac{X - M}{\sigma_{DM}} \quad \text{(see unit-I)}$$

or

$$Z = \frac{(M_1 \sim M_2) - (M_1 \sim M_2)}{\frac{\sqrt{\Sigma 1^2 + \sigma^2}}{N_1 + N_2}}$$

To distinguish the Z score of the difference of two sample means, the symbol C.R (Critical Ratio) is used. Therefore

$$C.R = \frac{(M_1 \sim M_2) - (M_1 \sim M_2)}{\Sigma_{DM}}$$

If the two independent populations are alike or same about a trait being measured, then

$$M_1 \sim M_2 = 0$$

$$\therefore \text{C.R.} = \frac{(M_1 \sim M_2) - 0}{\sigma_{DM}}$$

$$\text{Or} \quad \text{C.R} = \frac{(M_1 \sim M_2)}{\sigma_{DM}}$$

This is the general formula to decide the significance of the difference exists in the two sample means taken from the two independent populations.

The formula of C.R. clearly indicates that it is a simple ratio between difference of the two sample means and the standard error of the sample mean differences. Further it is nothing but a Z score, which indicates how far the two sample mean difference is deviating from the two parent population mean difference, which is Zero.

2.9.2 Levels of Significance

Whether a difference in the two statistic i.e. the statistical measures obtained or the parameters are to be considered as statistically significant?

It depends upon the probability that the given difference could have arisen "by chance.".

It also depends upon the purposes of experiment, usually, a difference is marked "significant", when the gap between the two sample means points to or signifies a real difference between parameters of the population from which the sample are drawn.

The research workers as the experimenters have an option to choose several arbitrary standards called levels of significance of which the .05 and .01 levels are most often used. The confidence with which an experimenter or research worker, rejects or retains (accept) a null hypothesis, depends upon the level of significance adopted.

You carefully look at the table Z distribution presented in unit 1, table no. 1.6.1, you will find that at the point ±1.96 the total 95% cases fall.

If we take ±1.96 at the base line of normal distribution curve as two points, we find that total 95% area of the curve lie between these two points.

Remaining 5% area lies left on to the right side of the curve i.e. 2½ % area lies to the left and 2 ½ % area lies to the right side of the curve.

If the Z Score of the mean difference which is also known as "C. R. value of t ratio" is 1.96 or below, it means the difference of the two means lies within the acceptance area of the normal distribution of the sampling distribution of the area differences.

Hence the null hypotheses (Ho) are retained. "CR" or "t ratio" is higher than 1.96, means the mean difference falls within the area of rejection, hence the null hypotheses (Ho) is rejected.

Further, you see the table 1.6.1 again, you will find that at the point ±2.58ó on the base of the normal distribution curve total 99% area of the curve or cases lie within the range -2.58 ó to + 2.58ó.

Only 1% area of the curve or cases lie beyond these two units. If the CR value or "t ratio" is below the value of 2.58, i.e. within the area of acceptance of 99% level the obtained mean difference is significant at .01 level or 99% level.

If the CR value of t. ratio is obtained above to the 2.58% , the null hypothesis (Ho) said to be rejected at 99% level or .01 level of significance.

2.9.3 The Null Hypothesis

In the above paragraphs a term Null Hypothesis is used in relation to determining of the significance of difference between the two means. Before we proceed further, it is essential to know about the null hypothesis and its rate in determining the significance of the difference of "Zero difference" or "No difference" in the relative specific parameters of the population and symbolically it is denoted as Ho.

Hypothesis is a suggested or pre determined relation of a problem which is tested on the basis of the evidences collected. Null hypothesis is a useful tool in testing the significance of the difference.

The null hypothesis a states that there is no true difference between two population means, and that the difference found between sample means, therefore, are only by chance i.e. accidental and unimportant.

The null hypothesis is based on the simple logic that a man is innocent until he is proved guilty. It constitutes or brings a challenge before the experimenter to call the necessary evidences to reject or retain the null hypothesis which he has framed.

After rejection of the null hypothesis automatically the alternative hypotheses will he accepted, For Example: In a study of Linguistic ability of boys and girls of group 14+years, the researcher has framed the following two hypothesis-

Ho: There is no difference in the means of linguistic ability scores of male and female adolescents of age group 14+ years.

$H_{A:}$ The mean of the linguistic ability scores is in favour of the adolescents girls than the boys of age group 14+ -16 + years.

It is obvious that if the null hypothesis (Ho) is rejected on the basis of statistical treatment made on the related evidences collected, the alternative hypothesis (H_A) will be accepted. If the null hypothesis (Ho) is accepted on the basis of the evidences collected, in such condition the alternative hypothesis (H_A) will be rejected.

2.9.4 Basic Assumption of Testing of Significance difference between the Two Sample Means

The formula which is used to test the significance of the difference between the two sample means (see 2.9.1) is based on certain basic assumptions. The assumptions are as under:

1) The variable or the trait being measured or studied is normally distributed in the universe.

2) There is no difference in the Means of the two or more populations. i.e. My=Mz If there is a violation or deviation in the above assumptions in testing the significant difference in the two means, we can not use "C.R" or "t" test of significance. In such condition, there are other methods, which are used for the purpose.

3) The samples are drawn from the population using random method of sample selection.

4) The size of the sample drawn from the population is relatively large.

2.9.5 Two Tailed and One Tailed Test of Significance

Under the null hypothesis, difference between the obtained sample means from two populations i.e. M_1 and M_2 may be either plus or minus and is often in one direction.

In yet another case, the differences between true parameters may be difference of Zero i.e. M_1-M_2=0, or Mz_{DM}=0, so that in determining probabilities we consider both traits of the sampling distribution. This two tailed test, as it is sometimes called, is generally used when we wish to discover whether two groups have conceivably been drawn from the same population with respect to the trait being measured.

In many research studies or experiments the primary concern is with the direction of the difference rather than with its existence in absolute terms. This situation arises when we are not interested in negative differences or in the losses made as this has no importance practically. However, we are much interested in the positive directions i.e. gains or growth or developments. For example, suppose we want to study the effect of extrinsic motivation on solving the mathematical problems or in sentence construction, it is unlikely that extrinsic motivation leads to loss in either solving the mathematical problems correctly or framing the sentences correctly.

Thus here only we are interested the positive effect of motivation and we study both gain made by the learners in solving the mathematical problems or constructing the sentences correctly. In such condition we use one tail of the normal probability curve that is the positive side and the Z or ó values will be changed after 95% and 99% level of significance far. In case of large samples the Z or ó values for 95% level it becomes 12.33 ó in place of 2.58ó.

2.9.6 Uncorrelated (Independent) and Correlated (Dependant) Sample Means

When we are interested to test whether two groups differ significantly on a trait or characteristics measured, the two situations arises with respect to differences between means

1) Uncorrelated or Independent two sample means

2) Correlated or Dependent two sample means

The two sample means are uncorrelated or independent when computed from different samples selected by using random method of sample selection from one population or from different populations or from uncorrelated tests administered to the same sample.

The two sample means are correlated when a single group of population is tested in two different situations by using the same test. In other words when one test is used on a single group before the experiment and after the experiment or when the units of the Group or the population from which two sample are drawn are not mutually exclusive.

In the latter situation, the modified formula for calculating the standard error of the difference of two sample means is applied.

Thus to test the significance of sample means, there are always following four s situations:

1) Two Large Independent Sample. i.e. when N_1 and N_2 >30

2) Two Small Independent Samples i.e. when N_1 and N_2 <30

3) Two Large correlated samples.

4) Two Small correlated samples.

2.10 SIGNIFICANCE OF THE TWO LARGE INDEPENDENT OR UNCORRELATED SAMPLE MEANS

The formula for testing the significance of two large independent sample means is as follows:

CR=M_1~M_2 where $ó_{DM} = \sqrt{\sigma M_1^2 + \sigma M_2^2}$ or

$= \sqrt{\sigma 1^2}$ / N1 + ó$2^{2/}$ N2

Self Assessment Questions

1) What do you mean by significance of the difference in two means?

.....

.....

.....

.....

2) Define Standard Error of the difference of the two sample means.

.....

.....

.....

.....

3) Define Sampling distribution of the differences of Means of two Samples.

.....

.....

.....

.....

4) What should be the mean value of sampling distribution of the difference of the means?

.....

.....

.....

.....

5) What indicates $S.E._{DM}$ or σ_{DM} ?

.....

.....

.....

.....

6) What do you mean by Ho, Define it.

.....

.....

.....

.....

7) What are the assumptions on which testing of the difference of two Mean is based?

.....

.....

.....

.....

8) What do you mean by One Tail Test and Two Tail Test? When these two tails are used?

.....

.....

.....

.....

9) What is meant by uncorrelated and correlated sample means?

.....

.....

.....

.....

Thus

CR = $M_1 \sim M_2$ / ó D

$$= \frac{M_1 \sim M_2}{\sqrt{\frac{\sigma 1^2}{N_1} + \frac{\sigma 2^2}{N_2}}}$$

Where

CR : Critical Ratio

M_1 : Mean of the Sample or Group 1

M_2 : Mean of the Sample or Group 2

$ó^1$: Standard Deviation of the Scores of sample 1

$ó^2$: Standard Deviation of the Scores of sample 2

N_1 : Number of cases in Sample 1

N_2 : Number of cases in sample 2

Example 8: An Intelligence test was administered on the two groups of Boys and Girls. These two groups were drawn from the two populations independently by using random method of sample selection. After administration of the test, the following statistics was obtained

Groups	N	M	Σ
Boys	65	52	13
Girls	60	48	12

Determine the difference between the mean values of Boys and Girls significant?

Solution

In the given problem, the two samples are quite large and independent. Therefore, to test the significance difference in the mean values of Boys and Girls. First we have to determine the null hypothesis which is

$Ho=M_B=M_G$ i.e.

There is no significant difference in the mean value of the Boys and Girls and the two groups are taken from the same population

$$\text{C.R.} = \frac{M_1 \sim M_2 - 0}{\sigma D}$$

$$= \frac{(M_1 \sim M_2)}{\sqrt{\frac{\sigma 1^2}{N_1} + \frac{\sigma 2^2}{N_2}}}$$

$$\text{C.R.} = \frac{(52-48)}{\sqrt{\frac{13^2}{65} + \frac{12^2}{60}}} = \frac{4}{\sqrt{\frac{169}{65} + \frac{144}{60}}} = \frac{4}{\sqrt{5}}$$

Or C.R. = 1.79

df = $(N_1-1) + (N_2-1)$

= (65-1) + (60-1)

= 123

To test the null hypothesis, which is framed, we will use two tail test. In the "t" distribution table (sub heading no. 2.5.2) at 123 df the "t" value at .05 level and .01 level is 1.98 and 2.62 respectively (The "t" table has 100 and 125 df, but df 123 is not given, therefore nearest of 123 i.e. 125df is considered). The obtained t value (1.79) is much less than these two values, hence it is not significant and null hypothesis is accepted at any level of significance.

Interpretation of the Results

Since our null hypothesis is retained, we can say that Boys and Girls do not differ significantly in their level of intelligence. Whatever difference is observed in the obtained mean values of two samples is due to chance factors and sampling fluctuations. Thus we can say with 99% level of confidence that no sex difference exists in the intelligence level of the population.

2.11 SIGNIFICANCE OF THE TWO SMALL INDEPENDENT ON UNCORRELATED SAMPLE MEANS

When the N's of two independent samples are small (less than 30), the S.E. $_{DM}$ or -ó$_{DM}$ (standard error of the difference between two means) should depend upon the Standard Deviation (S.D. or ó) values computed by using the correction

The formula used to test the significance of the difference of two means of small independent samples is :

$$t = \frac{M_1 \sim M_2}{S.E_{DM}}$$

Where

$$S.E_{DM} = S.D \frac{\sqrt{N_1 + N_2}}{N_1 \times N_2}$$

$$\text{And } S.D = \frac{\sqrt{\Sigma(x1\text{-}M1)^2 + \Sigma(x2\text{-}M2)^2}}{(N_1 - 1)(N_2 - 1)}$$

For simplification the above formula can also be written as

$$t = \frac{M1 \sim M2}{\frac{\Sigma d2 + \Sigma d2^2}{N1+N2-2} \times \frac{N1+N2}{N1N2}} \qquad \text{— (i)}$$

Where

$D_1 = (S_1\text{-}M_1)$, and

$D_2 = (x_2\text{-}M_2)$

Here, X_1 and X_2 are the new scores of two groups, M_1 and M_2 are given in relation to the two samples or groups having the small number units or cases.

When the raw data are not given and we have statistics or the estimates of two small size sample, in such condition, we use the formula-

The corresponding Mean values of the scores of the two groups N_1 and N_2 are the number of the units or cases in the two groups t is also a critical ratio in which more exact estimate of the $ó_{DM}$ is used. Here 't' in place C.R. is used because sampling distribution of "t" is not normal when N is small i.e. <30, "t" is a critical ratio (C.R.), but all C.R's are not "t's.

$$t = \frac{M_1 \sim M_2}{\sqrt{\sigma 1^2 (N_1 - 1)\sigma 2^2 (N_2 - 1) \times (N_1 + N_2) / N_1 N_2}} \quad \text{.....(ii)}$$

Where

M_1 = Mean of the scores of sample -1

M_2 = Mean of the scores of sample -2

$ó^1$ = Standard Deviation of the scores of sample-1

$ó^2$ = Standard Deviation of the scores of sample-2

N_1 = Number of units or cases on the sample-1

N_2 = Number of units or cases in the sample-2

For more clarification study the following examples very carefully.

Example 9: An attitude test regarding a vocational course was given to 10 urban boys and 5 rural boys. The obtained scores are as under-

Urban Boys (x1) = 6, 7, 8, 10, 15, 16, 9, 10, 0, 9

Rural Boys (x2) = 4, 3, 2, 1, 5

Determine at .05 level of significance that its there a significant difference in the attitude of boys belonging to rural and urban areas in relation to a vocational course?

Solution

Ho=b1=b2 : H1=b1 $\neq$ b2

Level of significance = .05

For acceptance or rejection of null hypothesis at .05 level of significance, the two tail test is used.

Thus

Urban Boys			**Rural Boys**		
X1	d1=(x1-m1)	d12	X2	d2=(X2-M2)	d22
6	-4	16	4	+1	1
7	-3	9	3	0	0
8	-2	4	2	-1	1
10	0	0	1	-2	4
15	+5	25	5	+2	4
16	+6	36			
9	-1	1	$\overline{\sum x_2=15}$		
10	0	0			$\overline{\sum d_2^2=10}$
10	0	0	M= 15/5		
9	-1	1	M= 3		

$\overline{\sum x_1=100}$ $\overline{\sum d1^2=92}$

M= ∑x/N
= 100/10
M=10

We know that

$$t = \frac{M_1 \sim M_2}{\sqrt{\frac{\Sigma d1^2+\Sigma d2^2}{N_1+N_2-2}\times\frac{N_1+N_2}{N_1N_2}}}$$

$$= \frac{10-3}{\sqrt{\frac{92+10}{10+5-2}\times\frac{10+5}{10\times5}}}$$

$$= \frac{7}{\sqrt{7.8\times0.30}} = \frac{7}{\sqrt{2.34}} = \frac{7}{1.54}$$

Or t = 4.46

df = $(N_1-1)+(N_2-1)$

= 9+4

=13

In "t" distribution table (table 2.5.1), the t value for 13 df at .05 level is 2.16. The obtained t value 4.46 is much greater than this value. Hence null hypothesis is rejected.

Interpretation of the Result

Our null hypothesis is rejected at .05 level of significance for 13 df. Thus we can say that in 95% cases significant difference in the attitude of the urban and rural boys regarding a vocational course. There are only 5% chances out of 100 that the two groups have same attitude towards a vocational course.

Example 10: music interest test was administered on 15 + years did boys and girls sample taken independently from the two populations. The following statistics was obtained:

	Mean	S.D.	N
Girls	40.39	8.69	30
Boys	35.81	8.33	25

Is the mean difference is in favour of girls?

Solution:

$Ho = b_1 = b_2$

$H_1 = b_1 \neq b_2$

In the given problem, the row scores of the two groups are not given. Therefore we will use the following formula for testing of the difference of means of two uncorrelated sample means:

$$t = \frac{M_1 \sim M_2}{\sqrt{\frac{\Sigma 1^2 (N_1 - 1) + \Sigma 2^2 (N_2 - 1)}{N_1 + N_2 - 2}} \times \frac{N_1 + N_2}{N_1 \times N_2}}$$

$$t = \frac{40.39 - 35.81}{\sqrt{\frac{(8.69)^2(30-1) + (8.33)^2(25-1)}{30+25-2}} \times \frac{30+25}{30 \times 25}}$$

$$= \frac{4.58}{\sqrt{75.516 \times 29 + 69.389 \times 24 \times 55}}$$

$$= \frac{4.58}{\sqrt{7274 \times .073}} = \frac{4.58}{2.309}$$

Or $t = 1.98$

$df = (N_1 - 1) + (N_2 - 1) = 53$

In the t distribution table for 53 df the t value at .05 level is 2.01. Our calculated t value 1.98 is less than this value. Therefore, the null hypothesis is retained.

Interpretation of the Results

Since our null hypothesis is accepted at .05 level of significance. Therefore it can be said that in 95 cases out of 100, there is no significant difference in the mean values of boys and girls regarding their interest in music. There are only 5% chances that the two groups do not have equal interest in music. Hence with 95% confidence, we can say that both boys and girls have equal interest in music. Whatever difference is deserved in the mean values of the groups is by chance or due to sampling of fluctuations.

2.12 SIGNIFICANCE OF THE TWO LARGE CORRELATED SAMPLES

In some of the experimental studies a single group is tested in two different conditions and the observations are in pairs. Or two groups are used in the experimental

condition, but they are matched by using pairs method. In these conditions, a modified formula of standard error of the difference of means is used. Therefore the formula for testing of the difference of two means of large correlated samples is –

$$t = \frac{M_1 \sim M_2}{\sqrt{\sigma M_1^2 + \sigma M_2^2 - 2r_{12}\sigma M_1\ \sigma M_2}}$$

In the formula

M_1 = Mean of the scores of sample -1

M_1 = Mean of the scores of sample -2

óM_1 = Standard Error of the Mean of sample-1

i.e. $\acute{o}M_1 = \frac{\Sigma 1}{\sqrt{N_1}}$

$\acute{o}M_2 = \frac{\Sigma 2}{\sqrt{N_1}}$

and r1,2 = correlation between two sets of scores.

For more classification go through the following examples

Example 11: An Intelligence test was administered on a group of 400 students twice after an interval of 2 months. The data obtained are as under-

		M	S.D
Testing –I	:	25	8
Testing-II	:	30	5
N	:	400	
r_{12}	:	65	

Test if there is a significant difference in the means of intelligence scores obtained in two testing conditions.

Solution:

Ho => b1=b2 and H1=>b1 ≠ b2

$$\therefore\ t = \frac{M_1 \sim M_2}{\sqrt{\sigma M_1^2 + \sigma M_2^2 - 2r_{12}\sigma M_1\ \sigma M_2}}$$

According the formula all values are given except S.E of means ($\sum M$). Therefore first we have to calculate standard errors of the means of the two sets of scores

$$\therefore\ \acute{o}M_1 = \frac{\Sigma 1^2}{\sqrt{N_1}} = \frac{8^2}{\sqrt{400}} = \frac{64}{20}$$

Or $\acute{o}M_1 = 3.20$

Similarly

$$\acute{o}M_2 = \frac{\sigma_2^{\,2}}{\sqrt{N_2}} = \frac{5^2}{\sqrt{400}} = \frac{25}{20}$$

Or $\acute{o}M_2 = 1.25$

Thus

$$t = \frac{30-25}{\sqrt{(3.20)^2 + (1.25)^2 - 2 \times .65 \times 3.20 \times 1.25}}$$

$$= \frac{5}{\sqrt{10.24 + 1.5625 - 5.20}}$$

$$= \frac{5}{\sqrt{6.6025}} = \frac{5}{2.57}$$

$t = 1.95$

df = N-1 = 400-1

df = N-1= 400-1 (In the example N is same i.e. the single group is tested in two different time intervals)

a df = 399

According to "t" distribution table (Table no-2.5.1) the value of t for 399 df at .01 level is 2.59. Our calculated value of t is 1.95, which is smaller than the value of t given in "t" distribution table. Hence the obtained t value is not significant even at .05 level. Therefore our null hypothesis is retained at .01 level of significance.

Interpretation of the Results

Since the obtained t value is found insignificant level for 399 df; thus the difference in the mean values of the intelligence scores of a group, tested after an interval of two months is not significant in 99 conditions out of 100, there is only 1% hence that the difference in two means is significant at .01 level.

Example 12: In a vocational training course an achievement test was administered on 64 students at the time of admission. After training of one year the same achievement test was administered. The results obtained are as under:

		M	ó
Before Training	:	52.50	7.25
After Training	:	58.70	5.30

Is the gain, after training significant?

Solution:

$H_o = b_1 = b_2$ (The gain after training is insignificant)

$H_1 = b_1 \neq b_2$

(*Note:* Read the problem carefully, here we will use one tail test rather to use two tail test. Because here we are interested in gain due to training, not in the loss. That is we are interested in one side of the B.P.C which is +ve side. 99% confidence and .05 for 95% confidence. See the table no-2.5.1 carefully and read the footnote)

We know that formula of testing the difference between two large correlated means is–

$$\text{Formula } t = \frac{M_1 \sim M_2}{\sqrt{\sigma M_1^{\,2} + \sigma M_2^{\,2} - 2r_{12}\sigma M_1\ \sigma M_2}}$$

Where

$$\acute{o}M_1 = \frac{\sigma 1}{\sqrt{N}} = \frac{7.25}{\sqrt{100}} = \frac{7.25}{10}$$

Or $\acute{o}M_1 = .725$

And $\acute{o}M_2$ $\frac{\sigma 2}{\sqrt{N}} = \frac{530}{\sqrt{N}} = \frac{5.30}{\sqrt{100}} = \frac{5.30}{10} = .53$

Or $\acute{o}M_2 = .53$

$$t = \frac{58.70 - 52.50}{\sqrt{(.725)^2 + (.53)^2 - 2 \times .50 \times .725 \times .53}}$$

$$= \frac{6.2}{\sqrt{0.4223}} = \frac{6.2}{.65}$$

$t = 9.54$

$df = (100\text{-}1)$

$= 99$

In the't' distribution table (table No. 2.5.1) at .02 level the t value for 99 df is 2.36 and out obtained t value is 9.54, which is much greater than the "t" value of the table. Thus the obtained t value is significant at 99% level of significance. Therefore our null hypothesis is rejected.

Interpretation of the Results

Since the obtained "t" value is found significant at .02 level for 99df. Thus we can say that gain on the achievement test made by the students after training is highly significant. Therefore we can say with 99% confidence that given vocational training is quite effective. There is only 1 chance out of hundred, the vocational training is ineffective.

2.13 SIGNIFICANCE OF TWO SMALL CORRELATED MEANS

In case of determining the significance difference between the two correlated small sample mean we have two methods, which are as under

i) **Direct Method:** i.e. we have to calculate Mean Values standard deviation values of the two groups and the coefficient of correlation (r_{12}) between the scores of two groups. In such condition the formula used to test significant difference in the

two means is –

$$t = \frac{M_1 \sim M_2}{\frac{\sqrt{\sigma 1^2}}{N-1} + \frac{\sqrt{\sigma 2^2}}{N-1} - 2r_{12}\frac{\sigma 1 \sigma 2}{N-1}}$$

$$t = \frac{M_1 \sim M_2}{\sqrt{Sm_1^2 + Sm_2^2 - 2r_{12}Sm_1Sm_2}}$$

Where

$Sm_1 = \frac{\sigma 1}{\sqrt{N-1}}$ (standard error of the small sample mean)

$Sm_2 = \frac{\sigma 2}{\sqrt{N-2}}$

ii) **Difference Method:** In this method we have the raw data of two small groups or sample and we are not calculate coefficient of correlation (r_{12}) between the two sets of scores.

Examples 13: A pre test and past test are given to 12 subjects. The scores obtained are as under–

S. No.-	1	2	3	4	5	6	7	8	9	10	11	12
Pre-Test:	42	50	51	26	35	42	60	41	70	38	62	55
Past-Test:	40	62	61	35	30	52	68	51	84	50	72	63

Determine if the gain on past test score significant?

Solution:

S.No. of Subjects	Post Test X_1	Pre Test X_2	Difference (X_2-X_1)	D-MD d	d^2
1	40	42	-2	-10	100
2	62	50	12	4	16
3	61	51	10	2	4
4	35	26	9	1	1
5	30	35	-5	-13	169
6	52	42	10	2	4
7	68	60	8	0	0
8	51	41	10	2	4
9	24	70	14	6	36
10	50	38	12	4	16
11	72	62	10	2	4
12	63	55	8	0	0
			$\sum D = 96$		$\sum d^2 = 354$

$MD = \frac{\sum D}{N}$ $SD = \sqrt{\frac{\sum d^2}{N-1}}$

$= \sqrt{\frac{354}{14}}$

Or SD= 5.67

$\therefore \quad SE_{DM} = \frac{\sigma D}{\sqrt{N}}$

Where

SE_{DM} = Standard Error of the Mean of Differences.

ΣD = Standard Deviation of the Differences

And N = Total No. of cases.

Thus $SE_{DM} = \frac{5.67}{\sqrt{12}} = \frac{5.67}{3.464}$

$= 1.631$

$\therefore\ t = \frac{MD}{SE_{DM}} = \frac{8}{1.637} = 4.88,\ df = 11$

In the "t" distribution table (Table 2.5.1 subheading 2.5.2) for 11 df at .02 level the value is 2.72 and our calculated value of t (4.88) is much greater than the table value. Therefore the null hypothesis is rejected at .01 level of significance.

Interpretation of the Results

Since our null hypothesis is rejected at.01 level of significance, therefore we can say that the gain made by the subject on past test is real in 99 cases out of 100. There are only 1% chance that the gain shown by the subjects in cue to chance factors as by sampling fluctuations.

2.14 POINTS TO BE REMEMBERED WHILE TESTING THE SIGNIFICANCE IN TWO MEANS

When you compare the means of two groups or to compare the means of a single group tested in two different situations or conditions and to know, whether the difference found in the two means is real and significant, or it is due to chance, factors, you should keep in mind the following steps as a process of testing the difference between two means.

i) Set up null hypothesis (Ho) and the alternative hypothesis (H1), according to the requirements of the problem.

ii) Decide about the level of significance for the test, usually in behavioural or social science, .05 and .01 levels are taken into consideration for acceptance or rejection of the null hypothesis.

iii) Decide whether one tailed or two tailed test of significance for independent or the correlated means.

iv) Decide whether the large or samples are involved in the problem or in the experiment.

v) Calculate either C.R value or "t" ratio value as per nature and size of the samples, by using the formulas discussed on the precious pages.

vi) Calculate degree of freedom (df). It should be N1+N2-2 for independent t or uncorrelated samples. While in case of correlated samples it should be N-1.

vii) Consult the "t" distribution table with df keeping in mind the level of significance, which we have decided at step number-11.

viii) Compare the calculated value of "t" with the "t" value given in the table with respect to df and level of significance.

ix) Interpret the Results:

If null hypothesis (Ho) is rejected, there is a significant difference between the two means.

If null hypothesis is accepted, there is no significant difference in the two means. Whatever the difference exists it has arisen due to sampling fluctuations or chance factors only.

Self Assessment Questions

1) How you can define "Critical Ratio" and "t" Ratio?

.....

.....

.....

.....

2) What is the difference between "CR" and "t"?

.....

.....

.....

.....

3) What is the difference between C.R. and Z Score?

.....

.....

.....

.....

4) How can you define the standard Error of the Difference of Means?

.....

.....

.....

.....

5) What formula you will use in the following conditions:

a) When two independent large samples are given.

.....

.....

.....

b) When two correlated large samples are given.

.....

.....

.....

c) When two independent small samples are given.

.....

.....

.....

d) When two small independent samples are given.

.....

.....

.....

6) What do you mean by independent samples?

.....

.....

.....

.....

7) What do you mean by correlated samples?

.....

.....

.....

.....

2.15 ERRORS IN THE INTERPRETATION OF THE RESULTS, WHILE TESTING THE SIGNIFICANT DIFFERENCE BETWEEN TWO MEANS

While interpreting the results obtained from the test of significance of a single mean or the difference between two means, we should take case and not to depend much on the statistical results obtained. Generally while interpreting the results, we may make following two type of Error.

Type-I Error or á Errors:

This Error occurs when the null hypothesis is true, but we reject the same by marking significant difference between the two means.

Type-II Error or â Error:

This error occurs when the null hypothesis is wrong or to be rejected while the same is retained.

This probability of occurrence of type-II error or â error due to finding high level of significance i.e. above to the .01 level of significance which may be .001 or the above.

2.16 LET US SUM UP

The Standard error of the estimates or statistics or sample statistical measures ($S.E._M$) consists –

Error of Sampling, and

Error of Measurement

Fluctuations from sample to sample, the so called sampling error or errors of sampling are not to be thought of as mistakes, features, and the like, but as variations arising from the fact that no two samples are over exactly alike.

Mean values and standard deviations (Σ's) obtained from random samples taken from a population are estimates of their parameters (the true statistical measurements of the population) and the standard error ($S.E_M$) measures the tepidness of these estimates.

If the $S.E_M$ is large, it does not follow necessarily that the obtained statistics is effected by a large sampling error, as much of the variations may be due to error of measurements, when error of measurements are low i.e. the measuring tools or tests are highly reliable, a large $S.E_M$ indicates considerable sampling error.

In the comparative studies or the experiments it is to decide whether the obtained differences of such magnitude is attributed to chance factor or sampling variations or it really exists? For such decisions the standard error of the difference is considered.

The critical or "t" ratio are nothing, but these are the Z scores , which tells how far the sample mean difference derivates to the population mean difference on a normal distribution curve.

"C.R" and "t" are the ratio of the mean difference of the two groups and the standard error of the mean differences.

While deciding the significance of any statistical measure or the difference on the means of two samples or two populations, the degree of freedom and levels of confidence are considered, and in the light of these two we either accept the null hypothesis or to reject the same.

While taking the decision a care is to be taken, so that type-I and type-II error should not be occur.

2.17 UNIT END QUESTIONS

1) Explain the term "Statistical Inference". How is the statistical inference is based upon the estimation of parameters.

2) Indicate the role of standard error for statistical generalisation.

3) Differentiate between significance of statistics and confidence interval of fiduciary limits.

4) Enumerate the various uses of Standard Error of the statistics.

5) What type of errors can occur while interpreting the results based on test of significance? How we can overcome these errors?

6) A Sample of 100 students with mean score 26.40 and SD 5.20 selected randomly from a population. Determine the .95 and .99 for confidence intervals for population true mean.

7) A small sample of 10 cases with mean score 175-50 and $\sum$ =5.82 selected randomly. Compute finding limits of parameter mean at .05 and .01 level of confidence.

8) The mean and standard deviation of the intelligence scores obtained on a group of 200 randomly selected students are 102 and 10.20 respectively. How dependable is mean I.Q. of the students?

The following are the data for two independent samples :

	N	M	S.D.
Boys	60	48.50	10.70
Girls	70	53.60	15.40

Is the difference in the mean values of Boys and Girls significant.

A reasoning ability test was given to 8 urban and 6 rural girls of the same Class. The data obtained are differ significantly in there reasoning ability.

Groups	Scores
Urban Girls	16,9,4,23,19,10,5,2
Rural Girls	20,5,1,16,2,4.

The observations given below obtained on 10 subjects in a experiment of Pre and Post test. Is gain trade by the students on post test significant?

Subjects	1	2	3	4	5	6	7	8	9	10
Scores on Pre Test	5	15	9	11	4	9	8	13	6	16
Scores on Post Test	7	9	4	15	6	13	9	5	6	12

9) A group of 10 students was given 5 trials on a test of physical efficiency. Their score on the I and V trials are given below. Test whether there is a significant effect of practice on the improvement made in first to fifth trial.

Subject	A	B	C	D	E	F	G	H	I	J
Trial I	15	16	17	20	25	30	17	18	10	12
Trial V	20	22	22	25	35	30	21	23	17	20

10) A group of 35 students randomly selected was tested before and after experimental treatment. The observations obtained are as under:

	M	ó
Pre Test	15.5	5.2
Post Test	21.6	4.8

Coefficient of Correlation between The scores of Pre and Post Test	0.70

Find out the groups is different significantly on the two testing conditions.

2.18 POINTS FOR DISCUSSION

1) What will happen on the standard error of sample mean if

   a) Sample is homogeneous and large

   b) Sample is heterogeneous and large

   c) Sample is heterogeneous and small

   d) Sample is homogeneous as well as small

2) Differentiate between "t" distribution and Z distribution. What is the basic difference between "t" Test and Z Test.

3) When the "t" distribution and "Z" distribution become conside

4) The necessity of a theoretical distribution model for estimation.

2.19 SUGGESTED READINGS

Aggarwal Y.P. (1990) *Statistical Methods – Concepts Application and Computation*. Sterling Publications Pvt Ltd. New Delhi.

Walker . H.M. and Lev. J. (1965). *Statistical Inference.* Oxford and I B H Publishing Co. Calcutta.

UNIT 3 ONE WAY ANALYSIS OF VARIANCE

Structure

3.0 Introduction

3.1 Objectives

3.2 Analysis of Variance

- 3.2.1 Meaning of the Variance
- 3.2.2 Characteristics of Variance
- 3.2.3 The Procedure of Analysis of Variance (ANOVA)
- 3.2.4 Steps of One Way Analysis of Variance
- 3.2.5 Assumptions Underlying Analysis of Variance
- 3.2.6 Relationship between F test and t test
- 3.2.7 Merits or Advantages of Analysis of Variance
- 3.2.8 Demerits or Limitations of Analysis of Variance

3.3 F Ratio Table and its Procedure to Use

3.4 Let Us Sum Up

3.5 Unit End Questions

3.6 Suggested Readings

3.0 INTRODUCTION

In the foregoing unit you have learned about how to test the significance of a mean obtained on the basis of observations taken from a group of persons and the test of significance of the differences between the two means. No doubt the test of significance of the difference between the two means is a very important technique of inferential statistics, which is used to test the null hypothesis scientifically and help to draw concrete conclusion. But its scope is very limited. It is only applicable to the two sets of scores or the scores obtained from two samples taken from a single population or from two different populations.

Now imagine if we have to compare the means of more than two populations or the number of groups, then what would happen? Can we apply successfully the Critical Ratio Test (CR) or the t test? The answer is yes, but not convenient to apply CR test or t test. The reason can be stated with an example. Suppose we have three groups A,B & C and we want to compare the significance difference in the means of the three groups, then first we have to make the pairs of groups e.g. A and B, then B and C, and then A and C and apply C.R. test or t test as the conditions required. In such condition we are to calculate three C.R. values or t values instead of one.

Now suppose we have eight groups and want to compare the difference in the means of the groups, in such condition we have to calculate 28 C.R. or t values as the condition may require.

It means when there are more than two groups say 3, 4, 5 ….. and k, it is not easy to apply 'C.R.' or 't' test of significance very conveniently.

Further 'C.R.' or 't' test of significance simply consider the means of two groups and test the significance of difference exists between the two means. It has no concern

in the variance that exist in the scores of the two groups or variance of the scores from the mean value of the groups.

For example let us say that A reaction time test was given to 5 boys and 5 girls of age group 15+ yrs. The scores were obtained in milliseconds are as given in the table below.

Groups	**Reaction time in M. Sec**					**Sum**	**Mean**
Girls	15	20	5	10	35	85	17M.Sec.
Boys	20	15	20	20	10	85	17M.Sec.

From the mean values shown in the table we can say that the two groups are equal in their reaction time and the average reaction time is 17 M. Sec. In this example, if we apply 't' test of significance, we will find, the difference in the two means insignificant and our null hypothesis is retained.

But if we look carefully to the individual scores of the reaction time of boys and girls, we will find that there is a difference in the two groups. The group of girls is very heterogeneous in their reaction time in comparison to the boys.

As the variation between the scores is ranging from 5 to 30 and deviation of scores from mean varies from 12 M. Sec. to 18 M. Sec.

While the group of boys is more homogeneous in their reaction time, as the variation in the individual scores is ranging from 5 to 10 and deviation of the scores from mean is 3 M. Sec to 7 M. Sec therefore group B is much better in their reaction time in comparison to the group A.

From, this example, you have seen that the test of significance of difference between the two means, some time lead us to draw wrong conclusion and we may wrongly retain the null hypotheses, though it should be rejected in real conditions.

Therefore, when we have more than two, say three or four or so forth and so on, the 'CR' or 't' test of significance are not very useful. In such condition, 'F' test is more suitable and it is known as one way analysis of variance. Because we are testing the significance difference in the average variance exists between the two or more than two groups, instead to test the significance of the difference of the means of the groups.

In this unit we will be dealing with F test or the analysis of variance.

3.1 OBJECTIVES

After going through this unit, you will be able to:

- Define variance;
- Differentiate between variance and standard deviation;
- Define analysis of variance;
- Explain when to use the analysis of variance;

- Describe the process of analysis of variance;
- Apply analysis of variance to obtain 'F' Ratio and to solve related problems;
- Analyse inferences after having the value of 'F' Ratio;
- Elucidate the assumptions of analysis of variance;
- List out the precautions while using analysis of variance; and
- consult the 'F' table correctly and interpret the results.

3.2 ANALYSIS OF VARIANCE

The analysis of variance is an important method for testing the variation observed in experimental situation into different part each part assignable to a known source, cause or factor.

In its simplest form, the analysis of variance is used to test the significance of the differences between the means of a number of different populations. The problem of testing the significance of the differences between the number of means results from experiments designed to study the variation in a dependent variable with variation in independent variable.

Thus the analysis of variance, as the name indicates, deals with variance rather than with standard deviations and standard errors. It is a method of dividing the variation observed in experimental data into different parts, each part assignable to a known source, cause or factor therefore

$$F = \frac{\text{Variance between the groups}}{\text{Variance within the groups}} = \frac{\sigma^2\text{Between the groups}}{\sigma^2\text{Within the groups}}$$

In which σ^2 is the population variance.

The technique of analysis of variance was first devised by Sir Ronald Fisher, an English statistician who is also known as the father of modern statistics as applied to social and behavioural sciences. It was first reported in 1923 and its early applications were in the field of agriculture. Since then it has found wide application in many areas of experimentation.

3.2.1 Meaning of the Variance

Before to go further the procedure and use of analysis of variance to test the significance difference between the means of various populations or groups at a time, it is very essential, first to have the clear concept of the term variance.

In the terminology of statistics the distance of scores from a central point i.e. Mean is called deviation and the index of variability is known as the mean deviations or standard deviation (σ).

In the study of sampling theory, some of the results may be some what more simply interpreted if the variance of a sample is defined as the sum of the squares of the deviation divided by its degree of freedom (N-1) rather than as the mean of the squares deviations.

The variance is the most important measure of variability of a group. It is simply the square of S.D. of the group, but its nature is quite different from standard deviation,

though formula for computing variance is same as standard deviation (S.D.)

$\therefore$ Variance = S.D.^{2} or $\sigma^2 = \frac{\Sigma(X-M)^2}{N}$

Where X : are the raw scores of a group, and

M : Mean of the raw scores.

Thus we can define variance as **"the average of sum of squares of deviation from the mean of the scores of a distribution."**

3.2.2 Characteristics of Variance

The following are the main features of variance:

- The variance is the measure of variability, which indicates the among groups or between groups difference as well as within group difference.
- The variance is always in plus sign.
- The variance is like an area. While S.D. has direction like length and breadth has the direction.
- The scores on normal curve are shown in terms of units, but variance is a area, therefore either it should be in left side or right side of the normal curve.
- The variance remain the some by adding or subtracting a constant in a set of data.

Self Assessment Questions

1) Define the term variance.

.....

.....

.....

.....

2) Enumerate the characteristics of Variance.

.....

.....

.....

.....

3) Differentiate between standard deviation and variance.

.....

.....

.....

.....

4) What do you mean by Analysis of Variance? Why it is preferred in comparison to 't' test while determining the significance difference in the means.

.....

.....

.....

.....

3.2.3 The Procedure of Analysis of Variance (ANOVA)

In its simplest form the analysis of variance can be used when two or more than two groups are compared on the basis of certain traits or characteristics or different treatments of simple independent variable is studied on a dependent variable and having two or more than two groups.

Before we discuss the procedure of analysis of variance, it is to be noted here that when we have taken a large group or a finite population, to represent its total units the symbol 'N' will be used.

When the large group is divided into two or more than two sub groups having equal number of units, the symbol 'n' will be used and for number of groups the symbol 'k' will be used.

Now, suppose in an experimental study, three randomly selected groups having equal number of units say 'N' have been assigned randomly, three kinds of reinforcement viz. verbal, kind and written were used. After a certain period, the achievement test was given to three groups and mean values of achievement scores were compared. The mean scores of three groups can then be compared by using ANOVA. Since there is only one factor i.e. type of reinforcement is involved, the situation warrants a single classification or one way ANOVA, and can be arranged as below:

Table 3.2.1

S.N.	Group - A		Group - B		Group – C	
	Scores of Verbal Reinforcement		Scores of Kind Reinforcement		Scores of Written Reinforcement	
	X_a	X_a^2	X_b	X_b^2	Xc_1	X_{c1}^2
	Xa_1	(X_{a1}^2)	X_{b1}	(X_{b1}^2)	Xc_1	(X_{c1}^2)
	Xa_2	(X_{a2}^2)	X_{b2}	(X_{b2}^2)	Xc_2	(X_{c2}^2)
	Xa_3	(X_{a3}^2)	X_{b3}	(X_{b3}^2)	Xc_3	(X_{c3}^2)
	Xa_4	(X_{a4}^2)	X_{b4}	(X_{b4}^2)	Xc_4	(X_{c4}^2)
	Xa_5	(X_{a5}^2)	X_{b5}	(X_{b5}^2)	Xc_5	(X_{c5}^2)
	.	.	.	.	.	.
	.	.	.	.	.	.
	Xa_n	(X_{an}^2)	X_{bn}	(X_{bn}^2)	Xc_n	(X_{cn}^2)
Sum	$\sum Xa$	$\sum X_a^2$	$\sum X_b$	$\sum X_b^2$	$\sum Xc1$	$\sum X_{c1}^2$
Mean	$\frac{\sum Xa}{n}$ =Ma		$\frac{\sum Xb}{n}$ =Mb		$\frac{\sum Xc}{n}$ =Mc	

To test the difference in the means i.e. MA, MB and MC, the one way analysis of variance is used. To apply one way analysis of variance, the following steps are to be followed:

Step 1 Correction tem $Cx = \frac{(\sum x)^2}{N} = \frac{(\sum x_a + \sum x_b + \sum x_x)^2}{n_1 + n_2 + n_3}$

Step 2 Sum of Squares of Total $SS_T = \sum x^2 - Cx$

$$= \sum x^2 - \frac{(\sum x)^2}{N}$$

$$= \left(\sum x_a^2 + \sum x_b^2 + \sum x_c^2\right) - \frac{(\sum x_3)}{N}$$

Step 3 Sum of Squares Among the Groups $SS_A = \frac{(\sum x)^2}{N} - Cx$

$$= \frac{(\sum x_a)^2}{n_1} + \frac{(\sum x_b)^2}{n_2} + \frac{(\sum x_c)^2}{n_3} - \frac{(\sum x)^2}{N}$$

Step 4 Sum of Squares Within the Groups $SS_W = SS_T - SS_A$

Step 5 Mean Scores of Squares Among the Groups $MSS_A = \frac{SS_A}{k-1}$

Where k = number of groups.

Step 6 Mean Sum of Squares Within the Groups $MSS_W = \frac{SS_W}{n-k}$

Where N = Total number of units.

Step 7 F Ratio i.e. $F = \frac{MSS_A}{MSS_W}$

Step 8 Summary of ANOVA

Table 3.2.2: Summary of ANOVA

Source of variance	Df	S.S.	M.S.S.	F Ratio
Among the Groups	k-1	SS_A	$\frac{SS_A}{K-1}$	$\frac{MSS_A}{MSS_W}$
Within the groups (Error Variance)	N-K	SS_W	$\frac{SS_w}{N-k}$	
Total	N-1			

The obtained F ratio in the summary table, furnishes a comprehensive or overall test of the significance of the difference among means of the groups. A significant F does not tell us which mean differ significantly from others.

If F-Ratio is not significant, the difference among means is insignificant. The existing or observed differences in the means is due to chance factors or some sampling fluctuations.

To decide whether obtain F-Ratio is significant or not we are taking the help of F table from a statistics book.

The obtained F-Ratio is compared with the F value given in the table keeping in mind two degrees of freedom k-1 which is also known as greater degree of freedom or df_1 and N-k, which is known as smaller degree of freedom or df_2. Thus, while testing the significance of the F ratio, two situations may arise.

The obtained F Ratio is Insignificant:

When the obtained F ratio is found less than the value of F ratio given in F table for corresponding lower degrees of freedom df_1 that is, k-1 and higher degree of freedom df that is, (df=N-K) (See F table in a Statistics Book) at .05 and .01 level of significance it is found to be significant or not significant. Thus the null hypothesis is rejected retained. There is no reason for further testing, as none of the mean difference will be significant.

When the obtained 'F Ratio' is found higher than the value of F ratio given in F table for its corresponding df_1 and df_2 at .05 level of .01 level, it is said to be significant. In such condition, we have to proceed further to test the separate differences among the two means, by applying 't' test of significance. This further procedure of testing significant difference among the two means is known as post-hoc test or post ANOVA test of difference.

To have clear understanding, go through the following working examples very carefully.

Example 1

In a study of intelligence, a group of 5 students of class IX studying each in Arts, Commerce and Science stream were selected by using random method of sample selection. An intelligence test was administered to them and the scores obtained are as under. Determine, whether the three groups differ in their level of intelligence.

Table 3.2.3

S.No.	Arts Group Intelligence scores	Comm. Group Intelligence scores	Science Group Intelligence scores
1	15	12	12
2	14	14	15
3	11	10	14
4	12	13	10
5	10	11	10

Solution: In the example k = 3 (i.e. 3 groups), n =5 (i.e. each group having 5 cases), n = 15 (i.e. the total number of units in the group)

Null hypothesis $H_0 = \mu_1 = \mu_2 = \mu_3$

i.e. the students of IX class studying in Arts, Commerce and Science stream do not differ in their level of intelligence.

Thus

Table 3.2.4

Arts Group		Commerce Group		Science Group	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
15	225	12	144	12	144
14	196	14	196	15	225
11	121	10	100	14	196
12	144	13	169	10	100
10	100	11	121	10	100
$\Sigma X_1 = 62$	$\Sigma X_1^2 = 786$	$\Sigma X_2 = 60$	$\Sigma X_2^2 = 730$	$\Sigma X_2 = 61$	$\Sigma X_3^2 = 765$
5		5		5	
12.40		12.00		12.20	

Step 1 : Correction term

$$Cx=\frac{\Sigma(x)^2}{N}=\frac{(\Sigma x_1+\Sigma x_2+\Sigma x_3.....\Sigma x_k)^2}{n_1+n_2+n_3.....n_k}=\frac{(62+62+61)^2}{5+5+5}=\frac{(183)^2}{15}$$

Or Cx = 2232.60

Step 2 : SS_T (Sum of squares of total) = $\Sigma x^2 - Cx$

Or $= \left(\Sigma x_1^2+\Sigma x_2^2+\Sigma x_3^2.....\Sigma x_k^2\right)-\frac{(\Sigma x)^2}{N}$

= (786+730+765) – 2232.60

= 2281.00 – 2232.60

SS_T = 48.40

Step 3 : SS_A (Sum of squares among the groups) = $\Sigma\frac{(\Sigma x)^2}{N} - Cx$

Or $= \frac{(\Sigma x_1)^2}{n_1}+\frac{(\Sigma x_2)^2}{n_2}+\frac{(\Sigma x_3)^2}{n_3}+.....+\frac{(\Sigma x_k)^2}{n_k}-Cx$

$= \frac{(62)^2}{5}+\frac{(60)^2}{5}+\frac{(61)^2}{5}-2232.60$

= 2233.00 – 2232.60

Or SS_A = 0.40

Step 4 : SS_W (Sum of squares within the groups) = $SS_T - SS_A$

Or = 48.40 – 0.40

SS_W = 48.00

Step 5 : MSS_A (Mean sum of squares among the groups)

$MSS_A = \frac{SSA}{k-1}=\frac{0.40}{3-1}=\frac{0.40}{2}$

Or MSS_A = 0.20

Step 6 : MSS_W (Mean sum of squares within the groups)

$$= \frac{SS_W}{N-K} = \frac{48}{15-3} = \frac{48}{12}$$

$MSS_W = 4.00$

Step 7 : F Ratio = $\frac{MSS_A}{MSS_W} = \frac{0.20}{4.00} = 0.05$

Step 8 : Summary of ANOVA

Table 3.2.5 : Summary of ANOVA

Source of variance	df	SS	MSS	F Ratio
Among the Groups	(k-1) 3-1 = 2	0.40	0.20	0.05
Within the Groups	(N-k) 15-3 = 12	48.00	4.00	
Total	14			

From F table (refer to statistics book) for 2 and 12 df at .05 level, the F value is 3.59. Our calculated F value is .05, which is very low than the F value given in the table. Therefore the obtained F ratio is not significant at .05 level of significance for 2 and 12 df. Thus the null hypothesis (H_0) is accepted.

Interpretation of Results

Because null hypothesis is rejected at .05 and .01 level of significance therefore with 99% confidence it can be said that the students studying in Arts, Commerce and Science stream do not differ significantly in their level of intelligence.

Example 2

An experimenter wanted to study the relative effects of four drugs on the physical growth of rats. The experimenter took a group of 20 rats of same age group, from same species and randomly divided them into four groups, having five rats in each group. The experimenter then gave 4 drops of corresponding drug as a one doze to each rat of the concerned group. The physical growth was measured in terms of weight. After one month treatment, the gain in weight is given below. Determine if the drugs are effective for physical growth? Find out if the drugs are equally effective and determine, which drug is more effective in comparison to other one.

Table 3.2.6 : Observations (Gain in weight in ounce)

Group A (Drug P)	Group B (Drug Q)	Group C (Drug R)	Group D (Drug S)
4	9	2	7
5	10	6	7
1	9	6	4
0	6	5	2
2	6	2	7

Solution: Given k = 4, n = 5, N = 20 and Scores of 20 rats in terms of weight

Null hypothesis $H_0 = \mu_1 = \mu_2 = \mu_3$

i.e. All the four drugs are equally effective for the physical growth of the rats.

Therefore:

Table 3.2.7

	Group A		Group B		Group C		Croup D	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
	4	16	9	81	2	4	7	49
	5	25	10	100	6	36	7	49
	1	1	9	81	6	36	4	36
	0	0	6	36	5	25	2	4
	2	4	6	36	2	4	7	49
Sum	12	46	40	334	21	105	27	167
n	5		5		5		5	
Mean	2.40		8.0		4.20		5.40	

Step 1 : Correction Term Cx = $\frac{(\sum x)^2}{N} = \frac{(12+40+21+27)^2}{20} = \frac{(100)^2}{20}$

= 500.00

Step 2 : Sum of Squares of total $SS_T = \sum x^2 - Cx$

= (46+334+105+167) – 500.00

= 152

Step 3 : Sum of Squares Among groups $SS_A = \sum \frac{(\sum x)^2}{n} - Cx$

$= \left(\frac{(12)^2}{5} + \frac{(40)^2}{5} + \frac{(21)^2}{5} + \frac{(27)^2}{5} \right) - 500.00$

= 82.80

Step 4 : Sum of Squares Within groups $SS_W = SS_T - SS_A$

= 152 – 82.80

= 69.20

Step 5 : Summary of ANOVA

Table 3.2.8: Summary of ANOVA

Source of variance	df	SS	MSS	F Ratio
Among Groups	4-1 = 3	82.80	$\frac{82.80}{3} = 27.60$	$\frac{27.60}{4.32} = 6.39$
Within Groups (Error variance)	40-4 = 16	69.20	$\frac{69.20}{16} = 4.32$	
Total	19			

In F table $F_{.05}$ for 3 and 16 df = 3.24

and

$F_{.01}$ for 3 and 16 df = 5.24

Our obtained F ratio (6.39) is greater than the F value at .01 level of significance for 3 and 16 df. Thus the obtained F ratio is significant at .01 level of confidence. Therefore the null hypothesis is rejected at .01 level of confidence. i.e. the drugs P, Q, R, S are not equally effective for physical growth.

In the given problem it is also to be determined which drug is comparatively more effective. Thus we have to make post-hoc comparisons.

For post-hoc comparisons, we apply 't' test of significance. The common formula of 't' test is –

$$t = \frac{M_1 \sim M_2}{S.E_{DM}}$$

Where :

M_1 = Mean of first group

M_2 = Mean of second group, and

SE_{DM} = Standard Error of Difference of Means.

Here $SE_{DM} = {}^{SDW}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Where SD_W or $\sigma_w = \sqrt{MSS_W}$

i.e. $S.D_W$ is the within groups S.D. and n_1 and n_2 are the size of the samples or groups being compared.

In the given example the means of four groups A, B, C and D are ranging from 2.40 ounce to 8.00 ounce, and the mean difference from 5.60 to 1.20. To determine the significance of the difference between any two selected means we must compute 't' ratio by dividing the given mean difference by its $S.E._{DM}$. The resulting t is then compared with the 't' value given in 't' table (Table no 2.5.1 of Unit 2) keeping in view the df of within the groups i.e. df_W. Thus in this way for four groups we have to calculate 6, 't' values as given below:

Step 6 : Standard deviation of within the groups

$SD_W = \sqrt{MSS_W} = \sqrt{4.32}$

= 2.08

Step 7 : Standard Error of Difference of Mean ($S.E_{DM}$)

$S.E._{DM} = {}^{SDW}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

= 1.31

(All the groups have same size therefore the value of SE_{DM} for the two groups will remain same)

Step 8 : Comparison of the means of the various pairs of groups.

Group A vs B

$$t = \frac{M_A - M_B}{S.E_{DM}} = \frac{8.0 - 2.40}{1.31} = \frac{5.60}{1.31} = 4.28$$ (Significant at .01 level for 16 df).

Group A vs C

$$t = \frac{4.20 - 2.40}{1.31} = \frac{1.80}{1.31} = 1.37$$ (Insignificant at .05 level for 16 df).

Group A vs D

$$t = \frac{5.40 - 2.40}{1.31} = \frac{3.0}{1.31}$$ t = 2.29 (Significant at .05 level for 16 df).

Group B vs C

$$t = \frac{8.0 - 4.90}{1.31} = \frac{3.80}{1.31}$$ t = 2.90 (Significant at .05 level for 16 df).

Group B vs D

$$t = \frac{8.0 - 5.40}{1.31} = \frac{2.60}{1.31} = 1.98$$ (Insignificant at .05 level for 16 df).

Group C vs D

$$t = \frac{5.40 - 4.20}{1.31} = \frac{1.20}{1.31} = 0.92$$ (Insignificant at .05 level for 16 df).

Results :

Out of 6 ‘t’ values, only 3 t values are found statistically significant. Among these three, one value is found significant at .01 level, while the two values are found significant at .05 level of significance. From these ‘t’ values, it is quite clear that the group B is better in physical growth in comparison to the group A and C, similarly group D is found better in comparison to Group A. The group B & D and groups C & D are found almost equal in their physical growth.

Interpretation of the Results

Since the group B is found better in physical growth of the rats in comparison to group A at 99% confidence level and at 95% confidence level it is found better in case of group C. But the group B and D are found approximately equally good in physical growth. Therefore the drug Q and S are effective for physical growth in comparison to the drugs P and R. Further the drug Q is comparatively more effective than the other drugs P, R and S respectively.

From the forgoing illustrations you have noted that if obtained F value is not significant, it means, there is no difference either of the pairs of groups. There is no need to follow ‘t’ test. If F is found significant, then the complete procedure of analysis of variance is to specify the findings by using ‘t’ test. Therefore you have noticed that only the F value is not sufficient when it is found significant. It is to be completed when supplemented by using the ‘t’ test.

3.2.4 Steps of One Way Analysis of Variance

From the foregoing two illustration, it is clear that following steps are to be followed when we use analysis of variance.

Step 1 : Set up null hypothesis.

Step 2 : Set the raw scores in table form as shown in the two illustrations.

Step 3 : Square the individual scores of all the sets and write the same in front of the corresponding raw score.

Step 4 : Obtain all the sum of raw scores and the squares of raw scores. Write them at the end of each column.

Step 5 : Obtain grand sums of raw scores as and square of raw square as $\sum x^2$

Step 6 : Calculate correction term by using the formula

$$\text{Cx} = \frac{\sum x^2}{N} \quad \text{Or} \quad \text{Cx} = \frac{\left(\sum x_1 + \sum x_2 + \sum x_3 ..... + \sum x_k\right)^2}{n_1 + n_2 + n_3 + ...n_k}$$

Step 7 : Calculate sum of squares i.e. SS_T by using the formula-

$$SS_T = \sum x^2 - Cx$$

Step 8 : Calculate sum of squares among the groups i.e. SS_A by using the formula-

$$SS_A = \frac{\sum x^2}{n} - Cx$$

$$\text{Or } SS_A = \frac{\left(\sum x_1^2\right)^2}{n_1} + \frac{\left(\sum x_2^2\right)^2}{n_2} + \frac{\left(\sum x_3^2\right)^2}{n_3} + ... + \frac{\left(\sum x_k^2\right)^2}{n_k} - Cx$$

Step 9 : Calculate sum of squares within the groups i.e. SSw by using the formula

$$SS_W = SS_T - SS_A$$

Step 10 : Calculate the degrees of freedom as

greater degree of freedom $df_1 = k - 1$ (where k is number of groups)

Smaller degree of freedom df_2 = N-k (where N is the total number in the group)

Step 11 : Find the value of Mean sum of squares of two variances as-

Mean sum of squares between the group $MSS_A = \frac{SS_A}{k-1}$

Mean sum of squares within the groups $MSS_W = \frac{SS_W}{N-K}$

Step 12 : Prepare summary table of analysis of variance as shown in 3.2.5 or 3.2.8.

Step 13 : Evaluate obtained F Ratio with the F ratio value given in F table (Table no. 3.3.1) keeping in mind df_1 and df_2.

Step 14 : Retain or Reject the Null Hypothesis framed as in step no-I.

Step 15 : If F ratio is found insignificant and null hypothesis is retained, stop further calculation, and interpret the results accordingly. If F ratio is found significant and null hypothesis is rejected, go for further calculations and use post-hoc comparison, find the t values and interpret the results accordingly.

3.2.5 Assumptions Underlying the Analysis of Variance

The method of analysis of variance has a number of assumption. The failure of the observations or data to satisfy these assumptions, leads to the invalid inferences. The following are the main assumptions of analysis of variance.

The distribution of the dependent variable in the population under study is normal.

There exists homogeneity of variance i.e. the variance in the different sets of scores do not differ beyond chance, in other words $\sigma_1 = \sigma_2 = \sigma_3 = ... = \sigma_k$.

The samples of different groups are selected from the population by using random method of sample selection.

There is no significant difference in the means of various samples or groups taken from a population.

3.2.6 Relationship between 'F' test and 't' test

The F test and t test are complementary to each other, because-

't' is followed when 'F' value is significant for the specification of inferences.

'F' test is followed, when 't' value is not found significant. Because within groups variance is not evaluated by 't' test. It evaluate only the difference between variance.

There is a fixed relation between 't' and 'F'. the F is the square of 't', while 't' is a square root of F.

F = t^2 *or* $t = \sqrt{F}$

3.2.7 Merits or Advantages of Analysis of Variance

The analysis of variance technique has the following advantages:

- It is the improved technique over the 't' test or 'z' test, it evaluates both types of variance 'between' and 'within'.
- This technique is used for ascertaining the difference among several groups or treatments at a time. It is an economical device.
- It can involve more than one variable in studying their main effects and interaction effects.
- In some of the experimental design e.g. simple random design and levels X treatment designs are based on one way analysis of variance.
- If 't' is not significant, F test must be followed, to analyse the difference between two means.

3.2.8 Demerits or Limitations of Analysis of Variance

The analysis of variance technique has following limitations also:

- We have seen that analysis of variance techniques is based on certain assumptions e.g. normality and homogeneity of the variances among the groups. The departure of the data from these assumptions may effect adversely on the inferences.
- The F value provides global findings of difference among groups, but it can not specify the inference. Therefore, for complete analysis of variance, the 't' test is followed for specifying the statistical inference.

- It is time consuming process and requires the knowledge and skills of arithmetical operations as well the high vision for interpretations of the results.
- For the use of 'F' test, the statistical table of 'F' value is essential without it results can not we interpreted.

3.3 F RATIO TABLE AND THE PROCEDURE TO USE

The significance of difference between two means is analysed by using 't' test or 'z' test as it has been discussed in earlier unit 2. The calculated or obtained t value is evaluated with the help of 't' distribution table. With df at .05 and .01 levels of significance. The df is the main base to locate the 't' values at different levels of significance given in the table.

In a similar way the calculated F ratio value is evaluated by using F table (refer to statistics book) by considering the degree of freedom between the groups and within the groups.

You observe the F table, carefully, you will find there are rows and columns. In the first row there are the degrees of freedom for larges variance i.e. for greater mean squares or between the variance. The first column of the table has also degree of freedom of smaller mean squares or the variance within the groups. Along with these two degree of freedom the F ratio values are given at .05 and .01 level of significance. The normal print values is the values at .05 level and the bold or dark print values are at .01 level of significance.

In the first illustrated example in the summary table the F ratio value is 0.05, df_1, is 2 and df_2 is 12. For evaluating the obtained F value with F value given in table df_1 = 2 is for row and df2 is for column. In the column you will find 12, proceed horizontally or row wise and stop in column 2, you will find F values 3.88, which is for .05 level of confidence and 6.93 (in dark bold print) which is meant for .01 level of confidence. Our calculated F value .05 is much less than these two values. Hence the F ratio is not significant also at .05 level. Thus the null hypothesis is retained.

Self Assessment Questions

1) State the assumptions of ANOVA.

.....

.....

.....

.....

2) What happens when these assumptions are violated?

.....

.....

.....

.....

3) Compare the 'F Ratio' test and 't Ratio' test in terms of their relative merits and demerits.

.....

.....

.....

.....

4) What is the mathematical relationship between F and t.

.....

.....

.....

.....

5) What is the relationship between S.D. and Mean sum of squares within the groups?

.....

.....

.....

.....

6) When the post ANOVA test of difference is applied?

.....

.....

.....

.....

7) How many degree of freedom are associated with the variation in the data for- A comparison of four means for independent samples each containing 10 cases?

.....

.....

.....

.....

8) A comparison of three groups selected independently each containing 15 units.

.....

.....

.....

.....

3.4 LET US SUM UP

Analysis of variance is used to test the significance of the difference between the means of a number of different populations say two or more than two.

Analysis of variance deals with variance rather to deal with means and their standard error of the difference exist between the means.

The variance is the most important measure of variability of a group. It is simply the square of S.D. of the group i.e. $v = \sigma^2$

The problem of testing the significance of the differences between the number of means results from experiments designed to study the variation in a dependent variable with variation in independent variable.

Analysis of variance is used when difference in the means of two or more groups is found insignificant.

There is a fixed relationship between 't' ratio and 'F' ratio. The relationship can be expressed as $F = t^2$ or $t = \sqrt{F}$.

While determining the significance of calculated or obtained ratio, we consider two types of degrees of freedom. One greater i.e. degree of freedom between the groups and second smaller i.e. degree of freedom within the groups.

3.5 UNIT END QUESTIONS

1) The four groups are given four treatments, each group consists of 5 subjects. At the end of treatment a test is administered, the obtained scores are given in the following table. Test significance of difference among four treatments.

Scores of the treatment

Group – A	Group – B	Group – C	Group – D
X1	X2	X3	X4
14	19	12	17
15	20	16	17
11	19	16	14
10	16	15	12
12	16	12	17

2) A Test Anxiety test was given to three groups of students of X class, classified as high achievers, average achievers and low achievers. The scores obtained on the test are shown below. Are the three groups differ in their test anxiety.

High-achievers	Average achievers	Low achievers
15	19	12
14	20	14
11	16	12
12	19	12

3) Apply ANOVA on the following sets of scores. Interpret your results.

Set-I	Set-II	Set-III
10	3	10
7	3	11
6	3	10
10	3	5
4	3	6
3	3	8
2	3	9
1	3	12
8	3	9
9	3	10

4) Summary of analysis of variance is given below:

Source of variance	Df	SS	MSS	F
Between sets	2	180	90.00	17.11
Within sets	27	142	5.26	
Total	29			

Interpret the result obtained.

Note Table F values are

$F_{.05}$ for 2 and 27 df = 3.35

$F_{.01}$ for 2 and 27 df = 5.49

5) Given the following statistics of the two groups obtained on a verbal reasoning test:

Group	N	M	σ
Boys	95	29.20	11.60
Girls	83	30.90	7.80

Calculate:

't' ratio for the two groups.

'F' ratio for the two groups.

What should be the degree of freedom for 't' ratio.

What should be the degrees of freedom for 'F' ratio.

Interpret the results obtained on 't' ratio and 'F' ratio.

6) Why it is necessary to fulfill the assumptions of 'F' test, before to apply analysis of variance.

7) Why the 'F' ratio test and 't' ratio tests are complementary to each other.

8) What should be the various problems of psychology and education. Where the ANOVA can be used successfully.

3.6 SUGGESTED READINGS

Aggarwal, Y.P. (1990). *Statistical Methods – Concept, Applications, and Computation*. New Delhi : Sterling Publishers Pvt. Ltd.

Ferguson, G.A. (1974). *Statistical Analysis in Psychology and Education.* New York : McGraw Hill Book Co.

Garret, H.E. & Woodwarth, R.S. (1969). *Statistics in Psychology and Education.* Bombay : Vakils, Feffer & Simons Pvt. Ltd.

Guilford, J.P. & Benjamin, F. (1973). *Fundamental Statistics in Psychology and Education.* New York : McGraw Hill Book Co.

Srivastava, A.B.L. & Sharma, K.K. (1974). *Elementary Statistics in Psychology and Education.* New Delhi : Sterling Publishers Pvt. Ltd.

Walker, H.M. & Lev. J. (1965). *Statistical Inference.* Calcutta : Oxford & I.B.H. Publishing Co.

UNIT 4 TWO WAY ANALYSIS OF VARIANCE

Structure

4.0 Introduction

4.1 Objectives

4.2 Two Way Analysis of Variance

4.3 Interactional Effect

4.4 Merits and Demerits of Two Way ANOVA

4.4.1 Merits of Two Way Analysis of Variance

4.4.2 Demerits or Limitationsof Two Way ANOVA

4.5 Let Us Sum up

4.6 Unit End Questions

4.7 Suggested Readings

4.0 INTRODUCTION

In the preceding unit 3 we have learned about the one way analysis of variance. In this technique, the effect of one independent or one type of treatment was studied on single dependent variable, by taking number of groups from a population or from different population heaving different characteristics. Generally, in one way analysis of variance simple random design is used.

Now, suppose we want to study the effect of two independent variables on a single dependent variable. Further suppose our aim is to study the independent effects of the independent variables as well as their combined or joint effect on the dependent variable. For example a medicine company has developed two types of drugs to get relief from smoking habit. The company wants to know:

1) The independent effect of drug A on smoking behaviour,

2) The independent effect of drug B on smoking behaviour, and

3) The joint or interactional effect of drug A and B i.e. A x B on the smoking behaviour.

Take another example, in a field experiment, a psychologist wants to study effect of type of families on the cognitive development of the children of the age group 3 to 5+ years of age in relation to their sex.

In this field experiment there are two independent variables viz. Type of Family and gender of the Children. The dependent variable is Cognitive Development.

Further the type of family variable has two levels i.e. joint families and nuclear families.

Similarly the gender variable has also two levels viz. boys and girls.

The experimenter wants to study the independent effects of type of family (Joint vs Nuclear) gender (Boys vs Girls) and the interactional effect i.e. joint effect of type of family and gender on the dependent variable viz. Cognitive Development.

Such type of studies related to field experiments or real experiments are known as factorial design of 2 × 2 which indicates there are two independent variables each having two levels.

Like wise there are several situations in which the effect of two or more than two independent variable is studied on a single dependent variable.

In such experimental studies, the one way analysis of variance is not applicable. We have to use two, three or four way of analysis of variance, which depends upon the number of independent variables and their number of levels.

4.1 OBJECTIVES

After completing this unit, you will be able to:

- Define two way analysis of variance;
- Use analysis of variance vertically or column wise and horizontally or row wise;
- Explain the independent effects of two or more than two variables having each two or more than two levels;
- Explain the term interaction effect;
- Analyse the interaction effect of two or more than two variables;
- Differentiate between one way analysis of variance and two way analysis of variance;
- Analyse problems related to field experiments and true experiments where factorial designs are used;
- Explain the interactional effect of two variables on dependent variables; and
- Explain variables graphically.

4.2 TWO WAY ANALYSIS OF VARIANCE

In two way analysis of variance, usually the two independent variables are taken simultaneously. It has two main effects and one interactional or joint effect on dependent variable. In such condition we have to use analysis of variance in two way i.e. vertically as well as horizontally or we have to use ANOVA, column and row wise. To give an example , suppose you are interested to study the intelligence i.e. I.Q. level of boys and girls studying in VIII class in relation to their level of socio economic status (S.E.S.). in such condition you have the following 3 x 2 design.(Refer to table 4.2.1)

Table 4.2.1: SES, Intelligence and Gender factors

Groups	Levels of S.E.S.			
	High	Average	Low	Total
Boys	M_{HB}	M_{AB}	M_{LB}	M_B
Girls	M_{HG}	M_{AG}	M_{LG}	M_G
Total	M_H	M_A	M_L	M

In the table above,

M : Mean of intelligence scores.

M_{HB}, M_{AB}, & M_{LB} : Mean of intelligence scores of boys belonging to different levels of S.E.S. i.e. High, Average & Low respectively.

M_{HG}, M_{AG}, & M_{LG} : Mean of intelligence scores of girls belonging to different levels of S.E.S. respectively.

M_H, M_A, & M_L : Mean of the intelligence scores of students belonging to different levels of S.E.S. respectively.

M_B, M_G : Mean of the intelligence scores of boys and girls respectively.

From the above 3 x 2 contingency table, it is clear, first you have to study the significant difference in the means column wise or vertically, i.e. to compare the intelligence level of the students belonging to different categories of socio-economic status (High, Average and Low).

Second you have to study the significant difference in the means row wise or horizontally, i.e. to compose the intelligence level of the boys and girls.

Then you have to study the interactional or joint effect of sex and socio-economic status on intelligence level i.e. we have to compare the significant difference in the cell means of columns and rows.

Obviously, you have more than two groups, and to study the independent as well as interaction effect of the two variables viz. socio-economic status and sex on dependent variable viz. intelligence in terms of I.Q., you have to use two way analysis of variance i.e. to apply analysis of variance column and row wise.

Therefore, in two way analysis of variance technique, the following type of effects are to be tested:

- Significance of the effect of A variable on D.V.
- Significance of the effect of B variable on A.V.
- Significance of the interaction effect of A x B variables on D.V.

In two way analysis of variance, the format of summary table after applying the analysis of variance is as under-

Table 4.2.2: Summary of two way ANOVA

Source of variance	df	SS	MS	F Ratio
Among the groups				
Between the group A	$k_a - 1$	SS_A	MSS_A	$F_1 = \frac{MSS_A}{MSS_W}$
Between the Group B	$k_b - 1$	SS_B	MSS_B	$F_2 = \frac{MSS_B}{MSS_W}$
Interrelation A x B	$(k_a - 1)(k_b - 1)$	SS_{AxB}	MSS_{AxB}	$F_3 = \frac{MSS_{A\times B}}{MSS_W}$
Within the Groups (Error variance)	$N - k_a - k_b$			
Total	N – 1			

For interpretation of the obtained F ratios, we have to evaluate each F ratio value with the F ratio given in F table (refer to statistics book) keeping in view the corresponding grater and smaller df and the level of confidence. There may be two possibilities.

All the obtained F ratios may be found insignificant even at .05 level. This shows that there is no independent (i.e. individual) as well as interaction (i.e. joint) effect of

the two independent variables on dependent variable. Hence null hypothesis will retain. There is no need to do further calculations.

All the three obtained F ratio's may be found significant either at .05 level of significance or at .01 level of significance. This shows that there is a significant independent (i.e. individual) as well as interactional (i.e. joint) effect of the independent variables on the dependent variable. Therefore the null hypothesis is rejected. In such condition if the two independent variables have more than two levels i.e. three or four, we have to go for further calculations and use post-hoc comparisons by finding out various 't' values by pairing the groups.

Similarly the significant interactional effect will also be studied further by applying 't' test of significance or by applying graphical method.

At least one or two obtained F ratio will be found significant either at .05 level of significance or at .01 level of significance. Thus the null hypothesis may partially be retained. In such condition too we have to do further calculations, by making post-hoc comparisons and use 't' test of significance, if the independent variables have more than two levels.

For more clarification, go through the following illustration carefully.

Example 1

A researcher wanted to study the effect of anxiety and types of personality (Extroverts and Introverts) on the academic achievement of the undergraduate students. For the purpose, he has taken a sample of 20 undergraduates by using random method of sample selection. He administered related test and found following observations in relation to the academic achievement of the students.

Level of Anxiety

	Groups	High anxiety	Low anxiety
Type of Personality	Extroverts	12 13 14 15 14	14 14 13 15 15
	Introverts	14 16 16 16 15	11 10 12 12 16

Determine the independent as well as interactional effect of anxiety and types of personality on the academic achievement of the undergraduates.

Solutions:

Given

Two independent variables

a) type of personality having two levels viz. extroverts and introverts

b) Anxiety it has also two level viz. high anxiety and low anxiety.

Dependent variable scores

Academic achievement scores.

Number of groups i.e. k = 4.

Number of units in each cell i.e. n = 5.

Total no. of units i.e. N = 20.

To find out :

Independent effect of type of personality and anxiety on the academic achievement of the students.

Interactional i.e. joint effect of anxiety and type of personality on academic achievement of the students.

Therefore

H_0 : "There is no significant effect of types of personality and level of anxiety on academic achievement."

For convenience, the given 2 x 2 table is rearranged as under:

Table 4.2.3

S.N.	Extroverts				Introverts			
	High Anxiety		Low Anxiety		High Anxiety		Low Anxiety	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
1	12	144	14	196	14	196	11	121
2	13	169	14	196	16	256	10	100
3	14	196	13	169	16	256	12	144
4	15	225	15	225	16	256	12	144
5	14	196	15	225	15	225	16	256
Sums	68	930	71	1011	77	1189	61	765
N	5		5		5		5	
M	13.60		14.20		15.40		12.20	

Step 1 : Correction Term Cx = $\frac{\left(\sum x^2\right)}{N}$

$= \frac{(68+71+77+61)^2}{20} = \frac{(277)^2}{20}$

= 3836.45

Step 2 : Sum of Squares of Total $SS_T = \sum x^2 - Cx$

= 930+1011+1189+765 – 3836.45

= 58.55

Step 3 : Sum of Squares Among the Groups

$$SS_A = \sum \frac{\left(\sum x\right)^2}{n} - Cx$$

$= \frac{(68)^2}{5} + \frac{(71)^2}{5} + \frac{(77)^2}{5} + \frac{(61)^2}{5} = 3836.45$

= 26.55

Step 4 : Sum of squares Between the A Groups (i.e. between types of personality)

$$SS_{BTP} = \frac{(\sum x_1 + \sum x_2)^2}{n_1 + n_2} + \frac{(\sum x_3 + \sum x_4)^2}{n_3 + n_4} - Cx$$

= 3836.50 – 3836.45

= .05

Step 5 : Sum of squares Between the B Groups (i.e. Between level of Anxiety)

$$SS_{Anx} = \frac{(\sum x_1 + \sum x_2)^2}{n_1 + n_2} + \frac{(\sum x_3 + \sum x_4)^2}{n_3 + n_4} - Cx$$

$$= \frac{(68+77)^2}{5+5} + \frac{(71+61)^2}{5+5} - 3836.45$$

= 8.45

Step 6 : Sum of squares of Interaction

$$SS_{AxB} = SS_A - SS_{BTP} - SS_{BAnx}$$

i.e. SS_{AxB} = Sum of squares Among the Groups – Sum of Squares Between Type of Personality – Sum of Squares Between Anxiety Levels.

$$SS_{AxB} = 26.55 - 0.05 - 8.45$$

$$= 18.05$$

Step 7 : Sum of Squares Within the Groups

$$SS_W = SS_T - SS_A - SS_B$$

$$= 58.55 - 26.55$$

$$= 32.00$$

Step 8 : Preparation of Result of Summary Table

Table 4.2.4 : Summary of Two-way ANOVA

Source of variance	df	Sum of Squares (SS)	Mean SS (MSS)	F Ratio
Among the Groups	(k-1) (4-1=3)	(26.55)	$\frac{SS_A}{df} = \frac{26.55}{3} = 8.85$	$\frac{8.85}{2} = 4.425$
Between the Groups- SS_{B_1} (Types of personality)	(k_1-1) 2 – 1 = 1	0.05	$\frac{SS_{B_1}}{df} = \frac{.05}{1} = .05$	$\frac{.05}{2} = .025$
SS_{B_2} (Anxiety levels)	(k_2-1) 2 – 1 = 1	8.45	$\frac{SS_{B_2}}{df} = \frac{8.45}{1} = 8.45$	$\frac{8.45}{2} = 4.225$
SS_{AxB}	(k_1-1)(k_2-1) 1 x 1 = 1	18.05	$\frac{SS_{B_2 \times SS_{B_1}}}{df} = \frac{18.05}{1} = 18.05$	$\frac{18.05}{2} = 9.025$
Within the Groups	(N-k) 20 – 4 = 16	32.00	$\frac{SS_W}{df} = \frac{32}{16} = 2$	
Total	19			

In the F table (refer to statistics book) for 1 and 6 df, the F value at .01 and .05 level are 8.86 and 4.60 respectively.

Our calculated F values for type of personality and anxiety are smaller than the table F value 4.60.

Therefore the obtained F ratio values are not significant even at .05 level of significance. Hence the null hypotheses is in relation to Type of Personality and Anxiety are retained.

In case of interaction effect the obtained F ratio value 9.025 is found higher than the F value given in table at .01 level of significance. Thus the F for interaction effect is significant at .01 level. Hence, null hypothesis for interaction effect is rejected.

Interpretation of the Results

Since our null hypotheses are accepted at .05 and .01 level of significance, for type of personality, therefore it can be said that there is no independent as well as interactional effect of Types of Personality and levels of Anxiety on the academic achievement of the students. In other words it can be said that the students who are either Extroverts or Introverts are equally good in their academic performance.

Similarly, the anxiety level of the students do not cause any significant variation in the academic achievement of the students.

But the students having different type of personality and have different level of anxiety, their academic achievement varies in 99% cases. From the mean values in the table 4.2.3 it is evident that the students who are Extroverts and have low level of anxiety are comparatively good in their academic achievement (M = 14.20).

In the case of Introverts those who have high level of anxiety are better in their academic achievement (M = 15.40) in comparison to others.

Example 2

In a study, effect of intelligence and sex on the mathematical creativity a group of 40 students (20 boys and 20 girls) was selected from a population of high school going students by using random method of sample selection. A test of intelligence and mathematics creativity was administrated to them. The observations obtained are given below. Determine the independent as well as interactional effect of sex and Intelligence on the mathematical creativity of the high school going students.

Table 4.2.5: Observations obtained on the mathematical creativity test

Groups	Boys	Girls
High Intelligent	15 15 15 12 13 15 16 16 16 20	14 13 13 15 15 13 13 14 15 14
Low Intelligent	15 14 12 13 15 14 15 14 13 12	10 12 10 13 13 10 11 12 10 10
Total units	20	20

Solution:

Given :

Two independent variables A- Sex, B- Intelligence. Each having 2 levels.

Dependent variable : Mathematical Creativity

Number of Groups k = 4

Number of cases in each group n = 10

Total number of units in the group N = 40

To find out : i) Independent effect of intelligence and sex on mathematical creativity.

ii) Interactional effect of intelligence and sex on mathematical creativity.

H_0 : There is no significant independent as well as interactional effect of Intelligence and Sex on the mathematical creativity of the students.

Therefore.

Table 4.2.6

Groups	Boys (A1)				Girls (A2)			
S.No.	High Intelligence (B1)		Low Intelligence (B2)		High Intelligence (B1)		Low Intelligence (B2)	
	X1	X2	X1	X2	X1	X2	X1	X2
1	15	225	15	225	14	196	10	100
2	15	225	14	196	13	169	12	144
3	15	225	12	144	13	169	10	100
4	12	144	13	169	15	225	13	169
5	13	169	15	225	15	225	13	169
6	15	225	14	196	13	169	10	100
7	16	256	15	225	13	169	11	121
8	16	256	14	196	14	196	12	144
9	16	256	13	169	15	225	10	100
10	20	400	12	144	14	196	10	100
Sum	153	2381	137	1889	139	1939	111	1247
n	10		10		10		10	
Mean	15.30		13.70		13.90		11.10	

Step 1 : Correction Term = Cx = $\frac{\sum(x)^2}{N} = \frac{(153+137+139+111)^2}{40}$

= 7290.00

Step 2 : Sum of Squares of total $SS_T = \sum x^2 - Cx$

= (2381+1889+1939+1247) – 7290

= 166.00

Step 3 : Sum of Squares Among groups $SS_A = \sum\frac{\sum(x)^2}{N} - Cx$

$$= \left(\frac{(153)^2}{10}+\frac{(137)^2}{10}+\frac{(139)^2}{10}+\frac{(111)^2}{10}\right)-7290.00$$

= 92

Step 4 : Sum of squares Between the Groups (Sex)

$$SS_{BSex} = \frac{(\sum x_1+\sum x_2)^2}{n_1+n_2}+\frac{(\sum x_3+\sum x_4)^2}{n_3+n_4}-Cx$$

$$= \frac{(153+137)^2}{20}+\frac{(139+111)^2}{20}-7290$$

= 40.00

Step 5 : Sum of squares Between the Groups (Intelligence)

$$SS_{BInt} = \frac{(\sum x_1+\sum x_2)^2}{n_1+n_2}+\frac{(\sum x_3+\sum x_4)^2}{n_3+n_4}-Cx$$

$$= \frac{(153+139)^2}{10+10}+\frac{(137+111)^2}{10+10}-7290.00$$

= 48.40

Step 6 : Sum of squares Between the Interactions (Sex x Intelligence)

$$SS_{BSex\times Int} = SS_A - SS_{BSex} - SS_{BInt}$$

$$= 3.60$$

Step 7 : Sum of Squares within the Groups

$$SS_W = SS_T - SS_A$$

$$= 166 - 92$$

$$= 74.00$$

Step 8 : Preparation of Summary Table / Result table

Table 4.2.7 : Summary of Analysis of Variance

Source of variance	df	Sum of Squares SS	Mean SS MSS	F Ratio
1) Among the Groups	(k-1) 4-1=3	(92)	(30.67)	(14.88)
2) Between the Groups i) $SS_{B_{Sex}}$	(k_1-1) 2-1=1	40.00	40.00	19.42
ii) $SS_{B_{Int.}}$	(k_2-1) 2-1=1	48.40	48.40	23.49
iii) $SS_{B_{Sex\times Int.}}$	1x1=1	3.60	3.60	1.75
3.Within the Groups (Error variance)	(N-k) 40-4=36	74.00	2.06	
Total	39			

From F table, the value of $F_{.05}$ for 1 and 36 df = 4.12 and $F_{.01}$ for 1 and 36 df = 7.42

Interpretation of the Results:

Independent Effects

Sex : From the ANOVA summary table the F ratio value for Sex is found 19.42, which is high in comparison to the F value given in F table for 1 and 36 df. Therefore F ratio for Sex variable is found significant at .01 level. Hence null hypothesis is rejected. In conclusion it can be said that in 99% cases, the boys are high in mathematical creativity in comparison to the girls. There are only 1% chance that the girls are better in mathematical creativity than the boys.

Intelligence: From the ANOVA summary table the F ratio value for intelligence is found 23.49, which is also significant at .01 level for 1 and 36 df. Thus the null hypothesis is rejected at .01 level of confidence.

Therefore, in 99% cases the high intelligent high school going students are high in their mathematical creativity in comparison to the low intelligent students. Only in 1

case out of 100, the low intelligent high school going students are high in mathematical creativity.

Interactional Effect

From the ANOVA summary table it is evident that the F ratio for interactional effect is found insignificant even at .05 level of significance for 1 and 36 df. Thus the null hypothesis is accepted.

Therefore, the joint effect of sex and intelligence do not cause any significant variation in the scores of mathematical creativity. In other words both boys and girls who are high in their intelligence are equally good in their mathematical creativity.

Similarly the low intelligent boys and girls also do not differ in their mathematical creativity. In the group of boys the high intelligent and low intelligent high school going students also do not differ in their mathematical creativity. Similarly in the group of girls, the high intelligent and low intelligent girls are also do not differ significantly in their mathematical creativity. This fact is also confirmed from the following Figure A and B.

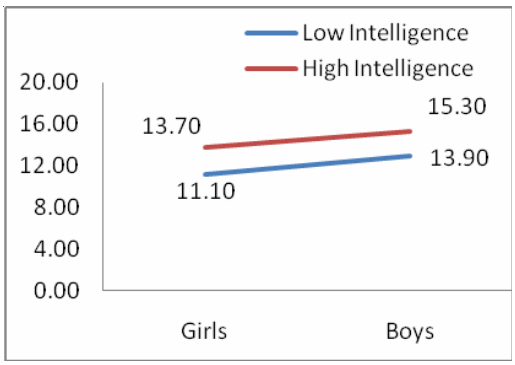


Fig. 4.2.1(A)

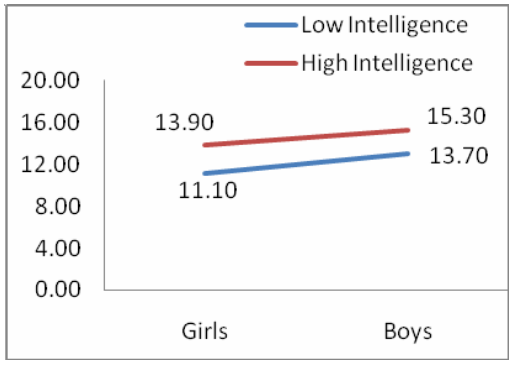


Fig. 4.2.1(B)

The two figures 4.2.1 (A) and 4.2.1 (B) both are showing two parallel lines. Which indicates that there is no interaction effect of sex and intelligence on the mathematical creativity of the high school going students.

Self Assessment Questions

1) What is the difference between one way analysis of variance and two way analysis of variance?

.....

.....

.....

.....

2) When we use two way analysis of variance?

.....

.....

.....

.....

3) In two way analysis of variance how many effects are tested.

.....

.....

.....

4) What indicates $K_{(a)}$, $K_{(b)}$ and $K_{(c)}$

.....

.....

.....

.....

5) What is meant by df_1 and df_2?

.....

.....

.....

.....

6) In what way we decide the significance of F ratio obtained in relation to various effects?

.....

.....

.....

.....

7) What do you mean by

2 x 2 Level design

3 x 3 Level design

2 x 4 Level design

3 x 3 Level design

.....

.....

.....

.....

4.3 INTERACTIONAL EFFECT

In the foregoing discussion we have frequently used the term "interaction" or "interactional effect" therefore it is essential to clarify the same.

In the two way analysis of variance, the consideration and interpretation of the interaction of variables or factors become important. Without considering the interaction between the different variables in a study, there is no use of two way or three way analysis of variance.

The interactions may be between two or more than two independent variables and its effect is measured on the dependent variable or the criterion variable. The need to know interaction effect on criterion variable or dependent variable is to know the combined effect of two or more than two independent variables on the criterion variable.

The reason, suppose there are two independent variables A & B and each has their own significance to create high variations in the criterion or dependent variable, but their joint or combined effect may cause very high, or low or their nullified effect on the dependent variable or criterion variable.

To have more clear idea let us suppose we have two types of fertilizers e.g. Urea and Phosphate. These have their own importance in the growth of the corps independently. But when we use the two chemical fertilizers in combined way with proper ratio, it might be possible, the growth of crops may be increased tremendously. Or it might be possible, the growth of crops may go down.

In the field of psychology and education suppose a treatment or a method of teaching A has its own significance to increase the level of achievement in a school subject, similarly the treatment B or a teaching method B is also good for encouraging results in academic achievement. But when we use the two methods of teaching jointly or give two treatments in combination we may get more encouraging results in academic achievement, or we do not have any significant effect on the achievement of the learners.

In the illustration-2, presented in this unit, compare the mean values, which we have shown in each cell in table 4.2.6. for convenience we are taking the same values here and presenting in the following 2 x 2 table.

Table 4.3.1

Intelligence	Boys M_1	Girls M_2	Total Mean
High	15.30	13.90	14.60
Low	13.70	11.10	12.40
Total Mean	14.50	12.50	13.50

In the above table, if we compare the total mean of first and second column, it is quite clear that there is a difference in the mean values of boys and girls and the higher mean is in the favour of boys. This is an independent effect of sex on mathematical creativity.

Similarly if we compare the total means of two rows we find, there is a difference in the means of high intelligent and low intelligent students and higher mean is in favour of the high intelligent group. It is actually the independent effect of intelligence on mathematical creativity.

Further in the above table 4.3.1, sex effects for boys and girls are

(14.50-13.50) = 1 and (12.50 – 13.50) = -1 respectively. If we subtract the first effect 1, from all averages in the first row and add 1 to all the averages in the second row, we have the following table:

Table 4.3.2: Sex factor

Groups	Boys	Girls	Total M
High Intelligence	14.30	12.90	13.60
Low Intelligence	14.70	12.10	13.40
Total Mean	14.50	12.50	13.50

Similarly in table 4.3.1 we subtract 1 from first column and add to the second column we have the following table:

Table 4.3.3: Sex factor

Groups	Boys	Girls	Total M
High Intelligence	14.30	12.90	14.60
Low Intelligence	14.70	12.10	12.40
Total M	13.50	13.50	13.50

Table 4.3.2 and table 4.3.3 give the intersectional resultant average, which show the direction of interaction and also indicates that there is no interaction effect of the A and B independent variable on dependent variable. In such condition if we plot the graph between the two independent variables we have approximately two parallel lines, as we have seen in the graphical presentation (see fig. 4.2.1 A and 4.2.1 B) respectively.

If there is a significant interactional effect of the two or more independent variables on the dependent variables; in such condition the graphical representation of the interactional effect will show two lines which are interacting at a point. For example, in example 1 the interactional effect of type of personality and anxiety is found significant at .01 level. If we draw the graph for interaction effect of Type of Personality and Level of Anxiety by considering the mean values of academic achievement, the obtained graph will be as under. (table 4.3.4. and graphs figures 4.3.1. A and B)

Table 4.3.4: the mean values of Extroverts and Introverts having high and low level of anxiety

Groups	Extroverts M1	Introverts M2	Total Mean
High Anxiety	13.60	15.40	14.50
Low Anxiety	14.20	12.20	13.20
Total Mean	13.90	13.80	13.85

(Mean values from table 4.2.3)

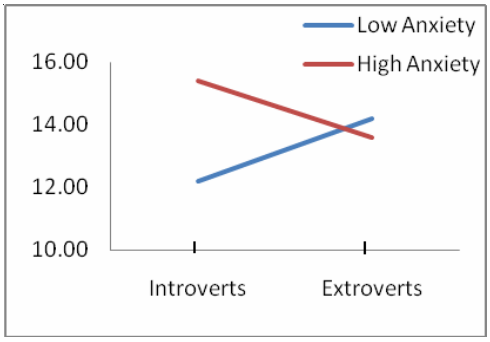


Fig. 4.3.1 (A)

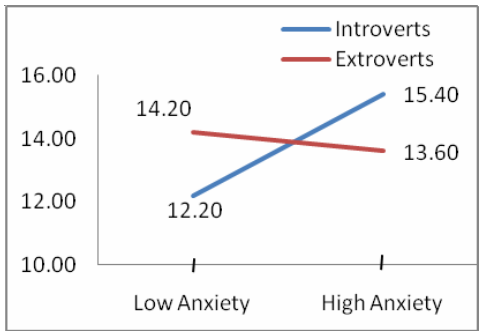


Fig. 4.3.1 (B)

4.4 MERITS AND DEMERITS OF TWO WAY ANOVA

4.4.1 Merits of Two Way Analysis of Variance

The following are the advantages of two way analysis of variance-

- This technique is used to analyse two types of effects viz. main effects and Interaction Effects.
- More than two factors effects are analysed by this technique.
- For analysing the data obtained on the basis of factorial designs, this technique is used.
- This technique is used to analyse the data for complex experimental studies.

4.4.2 Demerits or Limitations of Two Way ANOVA

The following limitations are found in this technique:

- When there are more than two classification of a factor or factors of study. F ratio value provides global picture of difference among the main treatment effects. The inference can be specified by using 't' test in case when F ratio is found significant for a treatment.
- This technique also follows the assumptions on which one way analysis of variance is based. If these assumptions are not fulfilled, the use of this technique may give us spurious results.
- This technique is difficult and time consuming.
- As the number of factors are increased in a study, the complexity of analysis in increased and interpretation of results become difficult.
- This technique requires high level arithmetical and calculative ability. Similarly it also requires high level of imaginative and logical ability to interpret the obtained results.

4.5 LET US SUM UP

The two way analysis of variance is a very important parametric technique of inferential statistics. It helps in taking concrete decisions about the effect of various treatments on criterion or dependent variable independently and jointly.

The independent effect of a variable on treatment means the direct or isolate effect of it on the dependent or criterion variable.

The interactional effect means joint effect of the two or more variables acting together on a dependent or criterion variables.

In case of insignificant interactional effect of the two or more predictors or independent variables on a criterion or dependable variable. The graphical representation will show the parallel lines, when the interactional effect is significant, its graphical representation will show the two crossed lines.

The two way analysis variance is a useful technique in experimental psychology as well as in experimental education especially in the field of teaching and learning. It is frequently used in field experiments or true experiments, when we use factorial designs specifically.

Interaction variance might be more reasonably expected in a combination of teacher and instruction method, of kind of task and method of attack by the learner, and of kind of reward when combined with a certain condition of motivation.

4.6 UNIT END QUESTIONS

1) What do you mean by Two-way Analysis of Variance.?

2) What is the difference between one way and two way ANOVA?

3) Indicate the graphical presentation of interaction effects?

4) Highlight the advantages and limitations of two way analysis of variance.

5) From the following hypothetical data, (Table below) determine-Which teaching method is effective than others. Also, Which teacher is contributing effectively in the learning outputs of the learners.

6) How far the joint effect of teaching method and the teacher is contributing in the learning performance of the students.

Teaching Methods

Teacher	A	B	C
T_1	10 7 6 10 4	3 3 3 3 3	10 11 10 5 6
T_2	3 1 8 9 2	3 3 3 3 3	8 9 12 9 10

Four groups of 8 students each having an equal number of boys and girls were selected randomly and assigned to different four conditions of an experiment. Test main effects due to conditions and sex and the interaction of the two conditions

Graphs	I	II	III	IV
Boys	7	9	12	12
	0	4	6	14
	5	5	10	9
	8	6	6	5
Girls	3	4	3	6
	3	7	7	7
	2	5	4	6
	0	2	6	5

7) In 4 × 3 factorial design 5 subjects are assigned randomly in each graph of 12 cells.

The following data obtained at the end of the experiment

Level of Intelligence	Method of Teaching			
	M1	M2	M3	M4
High (L_1)	6	8	7	9
	2	3	6	6
	4	7	9	8
	2	5	8	8
	6	2	5	9
Average (L_2)	4	6	9	7
	1	6	4	8
	5	2	8	4
	2	3	4	7
	3	6	8	4
Low (L_3)	4	3	6	6
	2	1	4	5
	1	1	3	7
	1	2	8	9
	2	3	4	8

Test the significance difference of difference of main effects and interaction effects.

4.7 SUGGESTED READINGS

Aggarwal, Y.P. (1990). *Statistical Methods-Concept, Applications, and Computation*. New Delhi : Sterling Publishers Pvt. Ltd.

Ferguson, G.A. (1974). *Statistical Analysis in Psychology and Education.* New York : McGraw Hill Book Co.

Garret, H.E. & Woodwarth, R.S. (1969). *Statistics in Psychology and Education.* Bombay : Vakils, Feffer & Simons Pvt. Ltd.

Guilford, J.P. & Benjamin, F. (1973). *Fundamental Statistics in Psychology and Education.* New York : McGraw Hill Book Co.

Srivastava, A.B.L. & Sharma, K.K. (1974). *Elementary Statistics in Psychology and Education.* New Delhi : Sterling Publishers Pvt. Ltd.

UNIT 1 RATIONALE FOR NON-PARAMETRIC STATISTICS

Structure

1.0 Introduction

1.1 Objectives

1.2 Definition of Non-parametric Statistics

1.3 Assumptions of Parametric and Non-parametric Statistics

- 1.3.1 Level of Measurement
- 1.3.2 Nominal Data
- 1.3.3 Ordinal Data
- 1.3.4 Interval and Ratio Data
- 1.3.5 Sample Size
- 1.3.6 Normality of the Data

1.4 The Use of Non-parametric Tests

- 1.4.1 Differences between Independent Groups
- 1.4.2 Differences between Dependent Groups
- 1.4.3 Relationships between Variables
- 1.4.4 Descriptive Statistics
- 1.4.5 Problems and Non-parametric Tests
- 1.4.6 Non-parametric Statistics
- 1.4.7 Advantages and Disadvantages of Non-parametric Statistics

1.5 Misconceptions about Non-parametric Tests

1.6 Let Us Sum Up

1.7 Unit End Questions

1.8 Suggested Readings and References

1.0 INTRODUCTION

Statistics is of great importance in the field of psychology. The human behaviour which is so unpredictable and cannot be so easily measured or quantified, through statistics attempts are made to quantify the same. The manner in which one could measure human behaviour is through normal distribution concept wherein it is assumed that most behaviours are by and large common to all and only a very small percentage is in either of the extremes of normal distribution curve. Keeping this as the frame of reference, the behaviour of the individual is seen and compared with this distribution. For analysis of obtained information about human behaviour we use both parametric and non-parametric statistics. Parametric statistics require normal distribution assumptions whereas non-parametric statistics does not require these assumptions and need not also be compared with normal curve. In this unit we will be dealing with non-parametric statistics, its role and functions and its typical characteristics and the various types of non-parametric statistics that can be used in the analysis of the data.

1.1 OBJECTIVES

After reading this unit, you will be able to:

- Define non-parametric statistics;
- Differentiate between parametric and non-parametric statistics;
- Elucidate the assumptions in non-parametric statistics;
- Describe the characteristics of non-parametric statistics; and
- Analyse the use of non-parametric statistics.

1.2 DEFINITION OF NON-PARAMETRIC STATISTICS

Non-parametric statistics covers techniques that do not rely on data belonging to any particular distribution. These include (i) distribution free methods (ii) non structural models. Distribution free means its interpretation does not depend on any parametrized distributions. It deals with statistics based on the ranks of observations and not necessarily on scores obtained by interval or ratio scales.

Non-parametric statistics is defined to be a function on a sample that has no dependency on a parameter. The interpretation does not depend on the population fitting any parametrized distributions. Statistics based on the ranks of observations are one example of such statistics and these play a central role in many non-parametric approaches.

Non-parametric techniques do not assume that the structure of a model is fixed. Typically, the model grows in size to accommodate the complexity of the data. In these techniques, individual variables are typically assumed to belong to parametric distributions, and assumptions about the types of connections among variables are also made.

Non-parametric methods are widely used for studying populations that are based on rank order (such as movie reviews receiving one to four stars). The use of non-parametric methods may be necessary when data have a ranking but no clear numerical interpretation. For instance when we try to assess preferences of the individuals, (e.g. I prefer Red more than White colour etc.), we use non parametic methods. Also when our data is based on measurement by ordinal scale, we use non-parametric statistics.

As non-parametric methods make fewer assumptions, their applicability is much wider than those of parametric methods. Another justification for the use of non-parametric methods is its simplicity. In certain cases, even when the use of parametric methods is justified, non-parametric methods may be easier to use. Due to the simplicity and greater robustness, non-parametric methods are seen by some statisticians as leaving less room for improper use and misunderstanding.

A statistic refers to the characteristics of a sample, such as the average score known as the mean. A parameter, on the other hand, refers to the characteristic of a population such as the average of a whole population. A statistic can be employed for either descriptive or inferential purposes and one can use either of the two types of statistical tests, viz., parametric Tests and non-parametric Tests (assumption free test).

The distinction employed between parametric and non-parametric test is primarily based on the level of measurement represented by the data that are being analysed. As a general rule, inferential statistical tests that evaluate categorical / nominal data and ordinal rank order data are categorised as non-parametric tests, while those tests that evaluate interval data or ratio data are categorised as parametric tests.

Differences between parametric and non-parametric statistics

The parametric and non-parametric statistics differ from each other on these various levels

Level of Differences	Parametric	Non Parametric
Assumed Distribution	Normal	Any
Assumed Variance	Homogeneous	Homogenous and Heterogeneous both
Typical data	Ratio or Interval	Ordinal or Nominal
Usual Central measure	Mean	Median
Benefits	Can Draw more conclusions	Simple and less affected by extreme score

Level of measurement is an important criterion to distinguish between the parametric and non-parametric tests. Its usage provides a reasonably simple and straightforward schema for categorisation that facilitates the decision making process for selecting an appropriate statistical test.

1.3 ASSUMPTIONS OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

Assumptions to be met for the use of parametric tests are given below:

- Normal distribution of the dependent variable
- A certain level of measurement: Interval data
- Adequate sample size (>30 recommended per group)
- An independence of observations, except with paired data
- Observations for the dependent variable have been randomly drawn
- Equal variance among sample populations
- Hypotheses usually made about numerical values, especially the mean

Assumptions of Non-parametric Statistics test are fewer than that of the parametric tests and these are given below.:

- An independence of observations, except with paired data
- Continuity of variable under study

Characteristics of non-parametric techniques:

- Fewer assumptions regarding the population distribution
- Sample sizes are often less stringent
- Measurement level may be nominal or ordinal
- Independence of randomly selected observations, except when paired
- Primary focus is on the rank ordering or frequencies of data
- Hypotheses are posed regarding ranks, medians, or frequencies of data

There are three major parametric assumptions, which are, and will continue to be routinely violated by research in psychology: level of measurement, sample size, and normal distribution of the dependent variable. The following sections will discuss these assumptions, and elucidate why much of the data procured in health science research violate these assumptions, thus implicating the use of non-parametric techniques.

Self Assessment Questions

1) Define Non-parametric statistics.

.....

.....

.....

.....

2) What are the charateristic featurs of non-parametric statistics?

.....

.....

.....

.....

3) Differentiate between parametric and non-parametric statistics.

.....

.....

.....

.....

4) What are the assumptions underlying parametric and non-parametric statistics?

.....

.....

.....

.....

When statistical tests are to be used one must know the following:

1.3.1 Level of Measurement

When deciding which statistical test to use, it is important to identify the level of measurement associated with the dependent variable of interest. Generally, for the use of a parametric test, a minimum of interval level measurement is required. Non-parametric techniques can be used with all levels of measurement, and are most frequently associated with nominal and ordinal level data.

1.3.2 Nominal Data

The first level of measurement is nominal, or categorical. Nominal scales are usually composed of two mutually exclusive named categories with no implied ordering: yes or no, male or female. Data are placed in one of the categories, and the numbers in each category are counted (also known as frequencies). The key to nominal level measurement is that there are no numerical values assigned to the variables. Given that no ordering or meaningful numerical distances between numbers exist in nominal measurement, we cannot obtain the coveted 'normal distribution' of the dependent variable. Descriptive

research in the health sciences would make use of the nominal scale often when collecting demographic data about target populations (i.e. pain present or not present, agree or disagree).

Example of an item using a nominal level measurement scale

1) Does your back problem affect your employment status? ☐ Yes ☐ No

2) Are you limited in how many minutes you are able to walk continuously with or without support (i.e. cane)? ☐ Yes ☐ No

1.3.3 Ordinal Data

The second level of measurement, which is also frequently associated with non-parametric statistics, is the ordinal scale (also known as rank-order). Ordinal level measurement gives us a quantitative 'order' of variables, in mutually exclusive categories, but no indication as to the value of the differences between the positions (squash ladders, army ranks). As such, the difference between positions in the ordered scale cannot be assumed to be equal. Examples of ordinal scales in health science research include pain scales, stress scales, and functional scales. One could estimate that someone with a score of 5 is in more pain, more stressed, or more functional than someone with a score of 3, but not by *how much*. There are a number of non-parametric techniques available to test hypotheses about differences between groups and relationships among variables, as well as descriptive statistics relying on rank ordering. Table below provides an example of an ordinal level item from the Oswestry Disability Index.

Table: Walking (Intensity of pain in terms of the ability to walk)

S. No.	Description	Intensity
1	Pain does not prevent me walking any distance	Lowest intensity of pain
2	Pain prevents me from walking more than 2 kilometres	Some level of intensity of pain
3	Pain prevents me from walking more than 1 kilometre	Moderate intensity of pain
4	Pain prevents me from walking more than 500 meters	High intensity of pain
5	I can only walk using a stick or crutches	Very high intensity of pain

1.3.4 Interval and Ratio Data

Interval level data is usually a minimum requirement for the use of parametric techniques. This type of data is also ordered into mutually exclusive categories, but in this case the divisions between categories are equidistant. The only difference between interval data and ratio data, is the presence of a meaningful zero point. In interval level measurement, zero does not represent the absence of value. As such, you cannot say that one point is two times larger than another. For example, 100 degrees Celsius is not two times hotter than 50 degrees because zero does not represent the complete absence of heat.

Ratio is the highest level of measurement and provides the most information. The level of measurement is characterised by equal intervals between variables, and a meaningful zero point. Examples of ratio level measurement include weight, blood pressure, and force. It is important to note that in health science research we often use multi item scales, with individual items being either nominal or ordinal.

1.3.5 Sample Size

Adequate sample size is another of the assumptions underlying parametric tests. In a large number of research studies, we do use small sample size and in certain cases we just use one case study and observe that case over a period of time. Some times, we take small sample sizes from a certain place and such samples are called as convenience samples, and limited funding. Thus, the assumption of large sample size is often violated by such studies using parametric statistical techniques.

The sample size required for a study has implications for both choices of statistical techniques and resulting power. It has been shown that sample size is directly related to researchers' ability to correctly reject a null hypothesis (power). As such, small sample sizes often reduce power and increase the chance of a type II error. It has been found that by using non-parametric techniques with small sample sizes, it is possible to gain adequate power. However, there does not seem to be a consensus among statisticians regarding what constitutes a small sample size. Many statisticians argue that if the sample size is very small, there may be no alternative to using a non-parametric statistical test, but the value of 'very small' is not delineated. It has been suggested by Wampold et al. (1990) , that the issue of sample size is closely related to the distribution of the dependent variable, given that as sample size increases, the sampling distribution approaches normal($n>100$).

At the same time, one can state that if the distribution of the dependant variable resembles closely the normal distribution, then it will amount to the sampling distribution of the mean being approximately normal. For other distributions, 30 observations might be required. Furthermore, in regard to decision about the statistical technique to be used, there is no clear cut choice but one can choose a technique depending on the nature of the data and sample size. Thus even the choice of parametric or non-parametric tests 'depends' on the nature of the data, the sample size, level of measurement, the researcher's knowledge of the variables' distribution in the population, and the shape of the distribution of the variable of interest. If in doubt, the researcher should try using both parametric and non-parametric techniques.

1.3.6 Normality of the Data

According to Pett (1997), in choosing a test we must consider the shape of the distribution of the variable of interest. In order to use a parametric test, we must assume a normal distribution of the dependent variable. However, in real research situations things do not come packaged with labels detailing the characteristics of the population of origin. Sometimes it is feasible to base assumptions of population distributions on empirical evidence, or past experience. However, often sample sizes are too small, or experience too limited to make any reasonable assumptions about the population parameters. Generally in practice, one is only able to say that a sample appears to come from say, a skewed, very peaked, or very flat population. Even when one has precise measurement (ratio scale), it may be irrational to assume a normal distribution, because this implies a certain degree of symmetry and spread.

Non-parametric statistics are designed to be used when we know nothing about the distribution of the variable of interest. Thus, we can apply non-parametric techniques to data from which the variable of interest does not belong to any specified distribution (i.e. normal distribution). Although there are many variables in existence that are normally distributed, such as weight, height and strength, this is not true of all variables in social or health sciences.

The incidence of rare disease and low prevalence conditions are both non-normally distributed populations. However, it seems that most researchers using parametric statistics often just 'assume' normality. Micceri et al. (1989) states that the naïve assumption of normality appears to characterise research in many fields.

However, empirical studies have documented non normal distributions in literature from a variety of fields. Micceri et al. (1989) investigated the distribution in 440 large sample achievement and psychometric measures. It was found that all of the samples were significantly non-normal ($p<0.01$).

It was concluded that the underlying tenets of normality assuming statistics appeared to be fallacious for the commonly used data in these samples. It is likely that if a similar study, investigating the nature of the distributions of data were to be conducted with some of the measures commonly used in health science research, a similar result would ensue, given that not all variables are normally distributed.

Self Assessment Questions

1) What are the aspects to be kept in mind before we decide to apply parametric or non-parametric tests?

   .....

   .....

   .....

2) What is ordinal data? Give suitable examples?

   .....

   .....

   .....

3) What are interval and ratio data ? Give examples.

   .....

   .....

   .....

4) Why is sample size important to decide about usingparametric or non-parametric tests?

   .....

   .....

   .....

5) What is meant by normality of a data? Explain.

   .....

   .....

   .....

1.4 THE USE OF NON-PARAMETRIC TESTS

It is apparent that there are a number of factors involved in choosing whether or not to use a non-parametric test, including level of measurement, sample size and sample distribution. When the choice of statistical technique for a set of data is not clear, there is no harm in analysing the data with both these methods, viz., parametric and on parametric methods.

It must be remembered that for each of the main parametric techniques there is a non-parametric test available. Also, experiments with the data would also determine which test provides the best power, and the greatest level of significance. In general, these tests fall into the following categories:

- Tests of differences between groups (independent samples);
- Tests of differences between variables (dependent samples);
- Tests of relationships between variables.

1.4.1 Differences between Independent Groups

Usually, when we have two samples that we want to compare concerning their mean value for some variable of interest, we would use the *t*-test for independent samples). The non-parametric alternatives for this test are the Wald-Wolfowitz runs test, the Mann-Whitney U test, and the Kolmogorov-Smirnov two-sample test.

If we have multiple groups, we would use analysis of variance (see ANOVA/MANOVA. The non-parametric equivalents to this method are the KruskalWallis analysis of ranks and the Median test.

1.4.2 Differences between Dependent Groups

If we want to compare two variables measured in the same sample we would customarily use the t-test for dependent samples. For example, we want to compare the math skills of students just at the beginning of the year and again at the end of the year, we would take the scores and use the t-test for such comparison and state that there is a significant difference between the two periods. Non-parametric alternatives to this test are the *Sign* test and *Wilcoxon's matched pairs* test.

If the variables of interest are dichotomous in nature (i.e., "pass" vs. "no pass") then McNemar's Chi-square test is appropriate.

If there are more than two variables that were measured in the same sample, then we would customarily use repeated measures ANOVA.

Non-parametric alternatives to this method are Friedman's two-way analysis of variance and Cochran Q test (if the variable was measured in terms of categories, e.g., "passed" vs. "failed"). Cochran Q is particularly useful for measuring changes in frequencies (proportions) across time.

1.4.3 Relationships between Variables

To express a relationship between two variables one usually computes the correlation coefficient. Non-parametric equivalents to the standard correlation coefficient of Pearson 'r' are Spearman R, Kendall Tau.

The appropriate non-parametric statistics for testing the relationship between two

variables are the Chi-square test, the Phi coefficient, and the Fisher exact test. In addition, a simultaneous test for relationships between multiple cases is available, as for example, Kendall coefficient of concordance. This test is often used for expressing inter rater agreement among independent judges who are rating (ranking) the same stimuli.

1.4.4 Descriptive Statistics

When one's data are not normally distributed, and the measurements at best contain rank order information, then using non prametric methods is the best. For example, in the area of psychometrics it is well known that the rated intensity of a stimulus (e.g., perceived brightness of a light) is often a logarithmic function of the actual intensity of the stimulus (brightness as measured in objective units of Lux). In this example, the simple mean rating (sum of ratings divided by the number of stimuli) is not an adequate summary of the average actual intensity of the stimuli. (In this example, one would probably rather compute the geometric mean.) Non-parametrics and Distributions will compute a wide variety of measures of location (mean, median, mode, etc.) and dispersion (variance, average deviation, quartile range, etc.) to provide the "complete picture" of one's data.

There are a number advantages in using non-parametric techniques in health science research. The most important of these advantages are the generality and wide scope of non-parametric techniques. The lack of stringent assumptions associated with non-parametric tests implies that there is little probability of violating assumptions, which implies robustness. The application of non-parametric tests in social and Health science researcsh is wide, given that they can be applied to constructs for which it is impossible to obtain quantitative measures (descriptive studies), as well as to small sample sizes.

1.4.5 Problems and Non-parametric Tests

The most common non-parametric tests used for four different problems include the following:

i) **Two or more independent groups:** The Mann-Whitney 'U' test and the Kruskal-Wallis one-way analysis of variance (H) provide tests of the null hypothesis that independent samples from two or more groups come from identical populations. Multiple comparisons are available by the Kruskal- Wallis test.

ii) **Paired observations:** The sign test and Wilcoxon Signed-rank test both test the hypothesis of no difference between paired observations.

iii) **Randomized blocks:** The Friedman two-way analysis of variance is the non-parametric equivalent of a two-way ANOVA with one observation per cell or a repeated measures design with a single group. Multiple comparisons are available for the Friedman test. Kendall's coefficient of concordanceis a normalization of the Friedman statistic.

iv) **Rank correlations:** The Kendall and Spearman rank correlations estimate the correlation between two variables based on the ranks of the observations.

The Table below gives an overview of when to use which test:

Choosing	**TEST**	
	PARAMETRIC	**NON PARAMETRIC**
Correlation test	Pearson	Spearman
Independent Measures, 2 Groups	Independent- Measures t-test	Mann-Whitney test ('U' Test)
Independent Measures, > 2 Groups	One Way Independent Measures ANOVA	Kruskal-Wallis Test
Repeated Measures, 2 Conditions	Matched-Pair t-Test	Wilcoxon test
Repeated Measures, > 2 Conditions	One-Way, Repeated Measures ANOVA	Friedman's Test

These statistics are discussed in many texts, including Siegel (1956), Hollander and Wolfe (1973), Conover (1980), and Lehmann (1975). Each of these non-parametric statistics has a parallel parametric test.

Self Assessment Questions

1) When do we use the non-parametric statistics?

.....

.....

.....

.....

2) What is meant by descriptive statistics in the context of non-parametric statistics?

.....

.....

.....

.....

3) State when to use which test – parametric or non-parametric?

.....

.....

.....

.....

4) What are the four problems for which non-parametric statistics is used?

.....

.....

.....

.....

1.4.6 Non-parametric Statistics

The primary barrier to use of non-parametric tests is the misconception that they are less powerful than their parametric counterparts (power is the ability to correctly reject the null hypothesis). It has been suggested that parametric tests are almost always more powerful than non-parametric tests. These assertions are often made with no references to support them, suggesting that this falls into the realm of 'common knowledge'. Evidence to support this is not abundant, nor conclusive. Rather, on closer examination, it is found that parametric tests are more powerful than non-parametric tests only if all of the assumptions underlying the parametric test are met.

Pierce (1970) suggests that unless it has been determined that the data do comply with all of the restrictions imposed by the parametric test; the greater power of the parametric test is irrelevant. This is because 'the purpose of applied statistics is to delineate and justify the inferences that can be made within the limits of existing knowledge - that purpose is defeated if the knowledge assumed is beyond that actually possessed'. Thus, the power advantage of the parametric test does not hold when the assumptions of the parametric test are not met, when the data are in ranks, or when the non-parametric test is used with interval or ratio data.

When comparison studies have been made between parametric and non-parametric tests, the non-parametric tests are frequently as powerful as parametric, especially with smaller sample sizes. Blair et al. (1985) compared the power of the paired sample t-test (a common parametric test), to the Wilcoxon signed-ranks test (non-parametric), under various population shapes and sample sizes (n=10, 25, 50), using a simple pre-post test design. It was found that in some situations the t-test was more powerful than the Wilcoxon.

However, the Wilcoxon test was found to be the more powerful test in a greater number of situations (certain population shapes and sample sizes), especially when sample sizes were small. In addition, the power advantage of the Wilcoxon test often increased with larger sample sizes, suggesting that non-parametric techniques need not be limited to studies with small sample sizes. It was concluded that insofar as these two statistics are concerned, the often-repeated claim that parametric tests are more powerful than non-parametric test is not justified.

Generally, the rationale for using the t-test over the Wilcoxon test is that the parametric tests are more powerful under the assumption of normality. However, it was shown in this study that even under normal theory, there was little to gain, in terms of power by using the t-test as opposed to the Wilcoxon.

It was suggested by Blair that 'it is difficult to justify the use of a t-test in situations where the shape of the sampled population is unknown on the basis that a power advantage will be gained if the populations does happen to be normal'. Blair concluded by saying that 'although there were only two tests compared here, it should be viewed as part of a small but growing body of evidence that is seriously challenging the traditional views of non-parametric statistics'. This study demonstrated that the use of non-parametric techniques is implicated whenever there is doubt regarding the fulfilment of parametric assumptions, such as normality or sample size.

Self Assessment Questions

Answer the following as True or False.

1) Parametric tests are equally assumptive as Non-parametric tests. T / F

2)	Non-parametric tests are most applicable when data is in rank form.	T / F
3)	Small sample size is not entertained by parametric tests.	T / F
4)	Parametric tests are more statistically grounded than Non-parametric tests.	T / F
5)	Non-parametric statistics cannot be used for complex research designs.	T / F

1.4.7 Advantages and Disadvantages of Non-parametric Statistics

Advantages

1) Non-parametric test make less stringent demands of the data. For standard parametric procedures to be valid, certain underlying conditions or assumptions must be met, particularly for smaller sample sizes. The one-sample t test, for example, requires that the observations be drawn from a normally distributed population. For two independent samples, the t test has the additional requirement that the population standard deviations be equal. If these assumptions/conditions are violated, the resulting P-values and confidence intervals may not be trustworthy. However, normality is not required for the Wilcoxon signed rank or rank sum tests to produce valid inferences about whether the median of a symmetric population is 0 or whether two samples are drawn from the same population.

2) Non-parametric procedures can sometimes be used to get a quick answer with little calculation.

   Two of the simplest non-parametric procedures are the sign test and median test. The *sign test* can be used with paired data to test the hypothesis that differences are equally likely to be positive or negative, (or, equivalently, that the median difference is 0). For small samples, an exact test of whether the proportion of positives is 0.5 can be obtained by using a binomial distribution. For large samples, the test statistic is:

   $(\text{plus} - \text{minus})^2 / (\text{plus} + \text{minus})$, where *plus* is the number of positive values and *minus* is the number of negative values. Under the null hypothesis that the positive and negative values are equally likely, the test statistic follows the chi-square distribution with 1 degree of freedom. Whether the sample size is small or large, the sign test provides a quick test of whether two paired treatments are equally effective simply by counting the number of times each treatment is better than the other.

   Example: 15 patients given both treatments A and B to test the hypothesis that they perform equally well. If 13 patients prefer A to B and 2 patients prefer B to A, the test statistic is $(13 - 2)^2 / (13 + 2)$ [= 8.07] with a corresponding P-value of 0.0045. The null hypothesis is therefore rejected.

   The *median test* is used to test whether two samples are drawn from populations with the same median. The median of the combined data set is calculated and each original observation is classified according to its original sample (A or B) and whether it is less than or greater than the overall median. The chi-square test for homogeneity of proportions in the resulting 2-by-2 table tests whether the population medians are equal.

3) Non-parametric methods provide an air of objectivity when there is no reliable (universally recognized) underlying scale for the original data and there is some

concern that the results of standard parametric techniques would be criticized for their dependence on an artificial metric. For example, patients might be asked whether they feel *extremely uncomfortable* / *uncomfortable* / *neutral* / *comfortable* / *very comfortable*. What scores should be assigned to the comfort categories and how do we know whether the outcome would change dramatically with a slight change in scoring? Some of these concerns are blunted when the data are converted to ranks[4].

4) A historical appeal of rank tests is that it was easy to construct tables of exact critical values, provided there were no ties in the data. The same critical value could be used for all data sets with the same number of observations because every data set is reduced to the ranks 1,...,*n*. However, this advantage has been eliminated by the ready availability of personal computers.

5) Sometimes the data do not constitute a random sample from a larger population. The data in hand are all there are. Standard parametric techniques based on sampling from larger populations are no longer appropriate. Because there are no larger populations, there are no population parameters to estimate. Nevertheless, certain kinds of non-parametric procedures can be applied to such data by using *randomization models*.

From Dallal (1988): Consider, for example, a situation in which a company's workers are assigned in haphazard fashion to work in one of two buildings. After a year physical tests are administered, it appears that workers in one building have higher lead levels in their blood. Standard sampling theory techniques are inappropriate because the workers do not represent samples from a large population—there is no large population. The randomization model, however, provides a means for carrying out statistical tests in such circumstances. The model states that if there were no influence exerted by the buildings, the lead levels of the workers in each building should be no different from what one would observe after combining all of the lead values into a single data set and dividing it in two, at random, according to the number of workers in each building. The stochastic component of the model, then, exists only in the analyst's head; it is not the result of some physical process, except insofar as the haphazard assignment of workers to buildings is truly random.

Of course, randomization tests cannot be applied blindly any more than normality can automatically be assumed when performing a t test. (Perhaps, in the lead levels example, one building's workers tend to live in urban settings while the other building's workers live in rural settings. Then the randomization model would be inappropriate.) Nevertheless, there will be many situations where the less stringent requirements of the randomization test will make it the test of choice. In the context of randomization models, randomization tests are the ONLY legitimate tests; standard parametric test are valid only as approximations to randomization tests.

Disadvantages

Such a strong case has been made for the benefits of non-parametric procedures that some might ask why parametric procedures are not abandoned entirely in favour of non-parametric methods!

The major disadvantage of non-parametric techniques is contained in its name. Because the procedures are *non-parametric*, there are no parameters to describe and it becomes more difficult to make quantitative statements about the actual difference between populations. (For example, when the sign test says two treatments are different, there is no confidence interval and the test does not say by how much the treatments differ.)

However, it is sometimes possible with the right software to compute estimates (and even confidence intervals!) for medians, differences between medians. However, the calculations are often too tedious for pencil-and-paper. A computer is required. As statistical software goes though its various iterations, such confidence intervals may become readily available, but are not there.

The second disadvantage is that non-parametric procedures throw away information. The sign test, for example, uses only the signs of the observations. Ranks preserve information about the order of the data but discard the actual values. Because information is discarded, non-parametric procedures can never be as powerful (able to detect existing differences) as their parametric counterparts when parametric tests can be used.

1.5 MISCONCEPTIONS ABOUT NON-PARAMETRIC TESTS

The lack of use of non-parametric techniques is owing to a series of common misconceptions about this branch of statistics.

Non-parametric statistics have long taken the back seat to parametric statistics, often being portrayed as inferior in practice and teaching. It has been suggested that researchers are hesitant to use these techniques, due to fears that peer reviewers may not be completely familiar with these statistics, and therefore unable to properly interpret, and review the results.

The above opinion could be due to the widespread case of limited exposure of researchers and clinicians to this type of statistics.

Non-parametric techniques are often left out of basic statistics courses, and relegated to the last chapter of texts, making them seem less important, while reinforcing the focus on parametric statistics.

Another common misconception concerning non-parametric statistics is that they are restricted in their application. It is thought that there are only a limited number of simple designs that can be analysed using these techniques.

However, there are non-parametric techniques which span from simple 2-group analysis, to complex structural equation modelling. Basically, for any parametric test, there is a non-parametric equivalent that would be equally, or in some cases, more appropriate for use.

Self Assessment Questions

1) What are the advantages of non-parametric statistics?

.....

.....

.....

2) What are the disadvantages of non-parametric statistics?

.....

.....

.....

3) What are the misconceptions about non-parametric statistic tests?

.....

.....

.....

1.6 LET US SUM UP

The key points of our discussion in this unit are:

1) **Characteristics common to most non-parametric techniques:**

- Fewer assumptions regarding the population distribution
- Sample sizes are often less stringent
- Measurement level may be nominal or ordinal
- Independence of randomly selected observations, except when paired
- Primary focus is on the rank ordering or frequencies of data
- Hypotheses are posed regarding ranks, medians, or frequencies of data

2) **Conditions when it is appropriate to use a non-parametric Test:**

- Nominal or ordinal level of measurement
- Small sample sizes
- Non-normal distribution of dependent variable
- Unequal variances across groups
- Data with notable outliers

3) **Advantages and disadvantages of Non-parametric Tests:**

- Methods quick and easy to apply
- Theory fairly simple
- Assumptions for tests easily satisfied
- Accommodate unusual or irregular sample distributions
- Basic data need not be actual measurements
- Use with small sample sizes
- Inherently robust due to lack of stringent assumptions
- Process of collecting data may conserve time and funds
- Often offer a selection of interchangeable methods
- Can be used with samples made up of observations from several different populations

1.7 UNIT END QUESTIONS

1) What are the major differences between parametric and non-parametric statistics?

2) Enumerate the advantages of non-parametric statistics.

3) Are there any assumptions for "Assumption Free tests"? If yes what are the assumptions of non-parametric statistics?

4) "Non-parametric Statistics has much wider scope than parametric statistics" support the statement with your arguments.

5) What are the major misconceptions regarding non-parametric statistics?

1.8 SUGGESTED READING

Cohen J. (1988) *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Micceri T. (1989) The Unicorn, The Normal Curve, and Other Improbable Creatures. *Psychological Bulletin*, 156-66.

Kerlinger F.N. (1964) *Foundations of Behavioural Research*. New York: Holt, Rinehart and Winston.

Pett M.A. (1997) *Non-parametric Statistics for Health Care Research*. London, Thousand Oaks, New Delhi: Sage Publications.

Siegel S. and Castellan N.J. (1988) *Non-parametric Statistics for the Behavioral Sciences* (2nd edition). New York: McGraw Hill.

Wampold BE & Drew CJ. (1990) *Theory and Application of Statistics*. New York: McGraw-Hill.

UNIT 2 MANN WHITNEY 'U' TEST FOR TWO SAMPLE TEST

Structure

2.0 Introduction

2.1 Objectives

2.2 Definition of Two Sample Tests

2.3 Mann Whitney 'U' Test

2.4 Relevant Background Information on 'U' Test

2.5 Step by Step Procedure for 'U' Test for Small Sample

2.6 Step by Step Procedure for 'U' Test for Large Sample

2.7 Computing Mann-Whitney U Test in SPSS

2.8 Wilcoxon Matched Pair Signed Rank Sum Test

2.9 Relevant Background Information on Wilcoxon Test

2.10 Step by Step Procedure for Wilcoxon Test for Small Sample

2.11 Step by Step Procedure for Wilcoxon Test for Large Sample

2.12 Computing the Wilcoxon Signed Rank Sum Test in SPSS

2.13 Comparison of Mann-Whitney 'U' Test and Wilcoxon MPSR test with *t*-test

2.14 Let Us Sum Up

2.15 Unit End Questions

2.16 Suggested Readings

2.0 INTRODUCTION

Non-parametric statistics are distribution free statistics and can be used for small samples as well as any kind of distribution. It has many tests which are equivalent to the parametric tests. For instance for the tests like mean, we have Mann Whitney U test, for Pearson 'r' we have Kendall tau test and so on. The non-parametric tests are available for single sample, matched pair sample, two samples and k samples. In this unit we will be dealing with Two sample tests and the various non-parametric tests that we can use analyse data if we have two samples. We will initially start with the definition of what is two sample test and go on to present different non-parametric statitistics that could be applied to analyse such data and then finally present how to solve problems based on such data.

2.1 OBJECTIVES

After reading this unit, you will be able to:

- Define two sample data;
- Explain what are two sample tests;

- Present the various non-parametric tests that can be used to analyse two sample data;
- Explain the significance levels and interpretation of such data; and
- Solve problems in two sample data.

2.2 DEFINITION OF TWO SAMPLE TESTS

Two sample tests are those which we call as tests of independence rather than goodness of fit tests. We are testing to see whether or not 2 variables are "related" or "dependent". Thus, the Ho , that is the Null Hypothesis, takes the general form

Ho: x and y are independent.

In a parametric test, we have seen earlier that to find out if two groups differ in performance , we used the t-test and if the t value was significant at .05 level, we rejected the null hypothesis and concluded that the two groups differed in their performance. In this type of t test we required the performance to be nornmally distributed and the sample size to be more than 30 and such other parametric test conditions. However if the sample size is less than 30 and the data is not normally distributed we would use the non-parametric test to find out if the two groups differed in their performance.

Let us say the 2 samples are males and females . We are comparing their marks in History at the final examination. Here one sample is gender, categorised into male and female. This is a nominal scale measurement. The other is 'scores obtained in History', a continuous variable which can range from zero to 100 or more depending on out of how many marks a score has been taken as performance.

Thus one variable , let us say Gender , that is X variable is in nominal scale of measurement and dichotomous (takes on only 1 of 2 possible values – that is male or female). The marks in History Y variable is treated as continuous (can take on a whole range of values on a continuum). To find out if males scored significantly higher than females in history in the final examination, we may if the sample size was more than 30, apply t-test and if the t value is significant, and also males average or the mean score is higher than that of female students, then we will conclude that males have scored significantly higher in history as compared to female students.

To take another example, let us say 2 groups matched in many respects each receive a different teaching method, and their final exam scores are compared.

Let us say X = Teaching method categorised into 1 and 2 methods.(X1 and X2)

Y = the final exam scores

Ho: X1 and X2 will not differ in terms of marks obtained

Here X = the independent variable and

Y = the dependent variable. (Marks in history)

We are trying to find out which of the two teaching methods is producing higher marks in final exam performance. The null hypothesis states there will be no difference in the marks obtained irrespective of X1 or X2 method of teaching.

Now all the subjects marks are taken for the two groups of persons undergoing two different methods of teaching. For this the ideal non-parametric test will be the t-test.

If the t-value goes beond the value given in the table at .05 level, we reject the null hypothesis and state that the two teaching methods do bring about a change in the performance of the students.

However if the sample size is small and there is no assumption of normality, then we would apply non-parametric test. These tests help in rejecting or accepting null hypothesis depending on the analysis.

Now let us see what are the various tests do we have under the non-parametrict test that can be applied.

2.3 MANN WHITNEY 'U' TEST

The Mann-Whitney (Wilcoxon) rank-sum test is a non-parametric analog of the two-sample *t* test for independent samples. The *Mann-Whitney U test* is a non-parametric test that can be used to analyse data from a two-group independent groups design when measurement is at least ordinal. It analyses the *degree of separation* (or the amount of overlap) between the Experimental (E) and Control (C) groups.

The *null hypothesis* assumes that the two sets of scores (E and C) are samples from the same population; and therefore, because sampling was random, the two sets of scores *do not differ systematically* from each other.

The *alternative hypothesis*, on the other hand, states that the two sets of scores *do* differ systematically. If the alternative is directional, or one-tailed, it further specifies the direction of the difference (i.e., Group E scores are systematically higher or lower than Group C scores).

The statistic that is calculated is either U or U'.

U1 = the number of Es less than Cs

U2 = the number of Cs less than Es

U = the smaller of the two values calculated above

U' = the larger of the two values calculated above

When you perform these tests, your data should consist of a random sample of observations from two different populations. Your goal is to compare either the location parameters (medians) or the scale parameters of the two populations. For example, suppose your data consist of the number of days in the hospital for two groups of patients: those who received a standard surgical procedure and those who received a new, experimental surgical procedure. These patients are a random sample from the population of patients who have received the two types of surgery. Your goal is to decide whether the median hospital stays differ for the two populations.

2.4 RELEVANT BACKGROUND INFORMATION ON 'U' TEST

The Mann-Whitney *U* test is employed with ordinal (rank-order) data in a hypothesis testing situation involving a design with two independent samples. If the result of the Mann-Whitney *U* test is significant, it indicates there is a significant difference between the two sample medians, and as a result of the latter the researcher can conclude there is a high likelihood that the samples represent populations with different median values.

Two versions of the test to be described under the label of the Mann-Whitney *U* test

were independently developed by Mann and Whitney (1947) and Wilcoxon (1949).

The version to be described here is commonly identified as the Mann-Whitney *U* test, while the version developed by Wilcoxon (1949) is usually referred to as the Wilcoxon-Mann-Whitney test.' Although they employ different equations and different tables, the two versions of the test yield comparable results.

In employing the Mann-Whitney *U* test, one of the following is true with regard to the rank order data that are evaluated:

a) The data are in a rank order format, since it is the only format in which scores are available; or

b) The data have been transformed into a rank order format from an interval ratio format, since the researcher has reason to believe that the normality assumption (as well as, perhaps, the homogeneity of variance assumption) of the t test for two independent samples (which is the parametric analog of the Mann-Whitney U test) is saliently violated.

It should be noted that when a researcher elects to transform a set of interval/ratio data into ranks, information is sacrificed. This latter fact accounts for the reluctance among some researchers to employ non-parametric tests such as the Mann-Whitney *U* test, even if there is reason to believe that one or more of the assumptions of the t test for two independent samples have been violated.

Various sources (e.g., Conover (1980, 1999), Daniel (1990), and Marascuilo and McSweeney (1977)) note that the Mann-Whitney U test is based on the following assumptions:

a) Each sample has been randomly selected from the population it represents;

b) The two samples are independent of one another;

c) The original variable observed (which is subsequently ranked) is a continuous random variable. In truth, this assumption, which is common to many non-parametric tests, is often not adhered to, in that such tests are often employed with a dependent variable which represents a discrete random variable; and

d) The underlying distributions from which the samples are derived are identical in shape. The shapes of the underlying population distributions, however, do not have to be normal.

Maxwell and Delaney (1990) pointed out the assumption of identically shaped distributions implies equal dispersion of data within each distribution. Because of this, they note that like the *t* test for two independent samples, the Mann-Whitney *U* test also assumes homogeneity of variance with respect to the underlying population distributions.

Because the latter assumption is not generally acknowledged for the Mann-Whitney *U* test, it is not uncommon for sources to state that violation of the homogeneity of variance assumption justifies use of the Mann-Whitney *U* test in lieu of the *t* test for two independent samples.

It should be pointed out, however, that there is some empirical evidence which suggests that the sampling distribution for the Mann-Whitney *U* test is not as affected by violation of the homogeneity of variance assumption as is the sampling distribution for *t* test for two independent samples. One reason cited by various sources for employing the Mann-

Whitney *U* test is that by virtue of ranking interval/ratio data, a researcher will be able to reduce or eliminate the impact of outliers.

Self Assessment Questions

1) Which non-parametric test should we use when the data is obtained from two different samples (Independent of each other) and we wish to see the difference between the two samples on a particular variable?

.....

.....

.....

.....

2) What is the underlying assumption of Mann-Whitney U test?

.....

.....

.....

.....

2.5 STEP BY STEP PROCEDURE FOR 'U' TEST FOR SMALL SAMPLE

Step-by-step procedure

Mann Whitney U Test for Small Sample case (not more than 20 items in each set), use U if the data is

a) in the form of ranks or

b) not normally distributed

c) there is an obvious difference in the variance of the two groups.

STEP 1: Rank the data (taking both groups together) giving rank 1 to the lowest score, and the highest rank to then highest score.

STEP 2: Find the sum of the ranks for the smaller sample

STEP 3: Find the sum of the ranks for the larger sample

STEP 4: Find U applying the formula given below:

$U = N_1N_2 + [N_1(N_1 + 1) / 2] - \Sigma R_1$

and

$U' = N_1N_2 + [N_2(N_2 + 1) / 2] - \Sigma R_2$

STEP 5: Look up the smaller of U and U' in Table H. There is a significant difference if the observed value is equal to or more than the table value.

STEP 6: Translate the results of the test back in the terms of experiment.

Worked Up Example:

Assessment Center Rating By Two Teams: Officers Randomly Assigned to Teams

Team A		Team B	
Score	Rank (R_1)	Score	Rank (R_2)
72	13	97	25
67	10	76	16
87	21	83	19
46	2	69	12
58	6	56	5
63	8	68	11
84	20	92	24
53	3	88	22
62	7	74	15
77	17	73	14
82	18	65	9
89	23	54	4
		43	1
	$\sum R1 = 148$		$\sum R2 = 177$

Step 1: Rank the ratings from lowest to highest regardless of assessment team.

Step 2: Sum the ranks in either group

$\Sigma (R_1) = 148$

$\Sigma (R_2) = 177$

Step 3: Calculate U

$U = N_1N_2 + [N_1(N_1 + 1) / 2] - \Sigma R_1$

$U = (12)(13) + [12(12 + 1) / 2] - 148$

$U = 156 + 78 – 148 = 86$

And Calculate U'

$U' = N_1N_2 + [N_2(N_2 + 1) / 2] - \Sigma R_2$

$U' = (12)(13) + [13(13 + 1) / 2] - 177$

$U' = 156 + 91 – 175 = 70$

Step 4: Determine the significance of U

Decide whether you are making a one- or a two-tailed decision

Compare the smaller value of U to the appropriate critical table value for N_1 and N_2

If the observed U is smaller than the table value, the result is significant.

Step 5: The critical value of U for $N_1 = 12$ and $N_2 = 13$, two-tailed $\alpha = 0.05$, is 41.

Since the smaller obtained value of U ($U' = 70$) is larger than the table value, the null hypothesis is accepted. And we conclude that there is no significant difference in the ratings given by the two assessment teams.

2.6 STEP BY STEP PROCEDURE FOR 'U' TEST FOR LARGE SAMPLE

When both sample sizes are greater than about 20, the sampling distribution of U is for practical purposes, normal. Therefore, under these conditions, one can perform a z-test as follows:

The procedure to obtain U is similar as in small sample case (Step 1 to 3). Then the formula for Z is applied as:

$$Z = [U - (N_1N_2) / 2] / \quad N_1N_2 \sqrt{(N_1 + N_2 + 1)/12}$$

If we are dealing with a two-tailed test, then the observed z is significant at the 5 per cent level if it exceeds 1.96. For one tailed test, 5 per cent significance is attained if z exceeds 1.64 (Check these in table D in Statistics book original).

The ranking procedure can become quite laborious in large samples. Partly for this reason and partly because violation of the assumptions behind parametric statistics become less important for large sample, the Mann Whitney U test tends to be restricted to use with relatively small samples.

Self Assessment Questions

1) What unit of sample is considered as an appropriate sample for Mann Whitney U test for small sample?

   .....

   .....

   .....

2) What is the rationale of applying Z test be applied in a non-parametric setting?

   .....

   .....

   .....

2.7 COMPUTING MANN-WHITNEY U TEST IN SPSS

Step 1.Choose Analyse

Step 2. Select Non-parametric Tests

Step 3. Select 2 Independent Samples

Step 4. Highlight your test variable (in our example this would be age) and click on the arrow to move this into the Test Variable List box

Step 5. Highlight the grouping variable and click on the arrow to move this into the Grouping Variable box.

Step 6. Click on Define Groups and type in the codes that indicate which group an observation belongs to (in our example, the codes which indicate whether a subject is male or female). Click on Continue

Step 7. Under Test Type make sure that Mann-Whitney U is selected

Step 8. If you want exact probabilities, click on Exact, choose Exact, then Continue

Click on OK

2.8 WILCOXON MATCHED PAIR SIGNED RANK TEST

The Wilcoxon Matched Pair signed-ranks test is a non-parametric test that can be used for 2 repeated (or correlated) measures when measurement is at least ordinal. But unlike the sign test, it *does* take into account the magnitude of the difference.

In using this test, the difference is obtained between each of N pairs of scores observed on matched objects, for example, the difference between pretest and post-test scores for a group of students.

The difference scores obtained are then ranked.

The ranks of negative score differences are summed and the ranks of positive score differences are summed.

The test statistic T is the smaller of these two sums.

Difference scores of 0 are eliminated since a rank cannot be assigned.

If the null hypothesis of no difference between the groups of scores is true, the sum of positive ranks should not differ from the sum of negative ranks beyond that expected by chance.

2.9 RELEVANT BACKGROUND INFORMATION ON WILCOXON TEST

The Wilcoxon matched-pairs signed-ranks test (Wilcoxon (1945, 1949)) is a non-parametric procedure employed in a hypothesis testing situation involving a design with two dependent samples. Whenever one or more of the assumptions of the t test for two dependent samples are saliently violated, the Wilcoxon matched-pairs signed-ranks test (which has less stringent assumptions) may be preferred as an alternative procedure.

The Wilcoxon matched-pairs signed-ranks test is essentially an extension of the Wilcoxon signed-ranks test (which is employed for a single sample design) to a design involving two dependent samples.

In order to employ the Wilcoxon matched-pairs signed ranks test, it is required that each of n subjects (or n pairs of matched subjects) has two interval/ratio scores (each score having been obtained under one of the two experimental conditions).

A difference score is computed for each subject (or pair of matched subjects) by subtracting a subject's score in Condition 2 from his score in Condition 1.

The hypothesis evaluated with the Wilcoxon matched-pairs signed-ranks test is whether or not in the underlying populations represented by the sampled experimental conditions, the median of the difference scores equals zero.

If a significant difference is obtained, it indicates that there is a high likelihood that the two sampled conditions represent two different populations.

The Wilcoxon matched-pairs signed-ranks test is based on the following assumptions:

a) The sample of n subjects has been randomly selected from the population it represents;

b) The original scores obtained for each of the subjects are in the format of interval/ratio data; and

c) The distribution of the difference scores in the populations represented by the two samples is symmetric about the median of the population of difference scores.

As is the case for the t test for two dependent samples, in order for the Wilcoxon matched pairs signed ranks test to generate valid results, the following guidelines should be adhered to:

a) To control for order effects, the presentation of the two experimental conditions should be random or, if appropriate, be counterbalanced; and

b) If matched samples are employed, within each pair of matched subjects each of the subjects should be randomly assigned to one of the two experimental conditions

As is the case with the t test for two dependent samples, the Wilcoxon matched-pairs signed-ranks test can also be employed to evaluate a "one-group pretest-posttest" design. The limitations of the one group pretest posttest design are also applicable when it is evaluated with the Wilcoxon matched pairs signed ranks test.

It should be noted that all of the other tests in this text that rank data (with the exception of the Wilcoxon signed-ranks test), ranks the original interval/ratio scores of subjects.

The Wilcoxon matched-pairs signed-ranks test, however, does not rank the original interval/ratio scores, but instead ranks the interval/ratio difference scores of subjects (or matched pairs of subjects).

For this reason, some sources categorise the Wilcoxon matched-pairs signed-ranks test as a test of interval/ratio data.

Most sources, however, categorise the Wilcoxon matched-pairs signed-ranks test as a test of ordinal data, by virtue of the fact that a ranking procedure is part of the test protocol.

Self Assessment Questions

1) Which non-parametric test should we use when the data is obtained from two related sample and we wish to see the difference between the two samples on a particular variable?

.....

.....

.....

.....

2) Which one assumption does not apply to Wilcoxon Matched Pair Test, which applies to Mann Whitney U test?

.....

.....

.....

.....

3) What is the difference between t Test for Matched Pair sample and Wilcoxon Matched Pair Test?

.....

.....

.....

.....

2.10 STEP BY STEP PROCEDURE FOR WILCOXON TEST FOR SMALL SAMPLE

Step-by-step procedure

Wilcoxon Test-Small Sample case (not more than 25 pairs of scores).

For matched pairs or repeated measures designs: use instead of a correlated *t*-test if either

a) the differences between treatments can only be ranked in size or

b) the data is obviously non-normal or

c) there is an obvious difference in the variance of the two groups.

STEP 1: Obtain the difference between each pair of reading, taking sign into account

STEP 2: Rank order these differences (ignoring the sign), giving rank 1 to the smallest difference

STEP 3: Obtain *T,* the sum of the ranks for differences with the less frequent sign

STEP 4: Consult Table J. If the observed *T* is equal to or less than the table value then there is a significant difference between two conditions

STEP 5: Translate the result of the test back in terms of the experiment

Worked Up Example:

Eight pairs of twins were tested in complex reaction time situations; one member of each pair was tested after drinking 3 double whiskies, the other member was completely sober. The following reaction times were recorded:

Sober Group	Whisky Group	Step 1: Differences	Step 2:Ranks
310	300	-10	1
340	320	-20	2
290	360	70	5
270	320	50	4
370	540	170	6
330	360	30	3
320	680	360	7
320	1120	800	8

STEP 3: Less frequent sign of difference is negative,

$T = 1 + 2 = 3$

STEP 4: From Table J, when N = 8, T = 4. As the observed value of *T* is less than the table value, there is a significant difference between the two conditions.

STEP 5: Complex reaction time scores are significantly higher after drinking 3 double whiskies than when sober.

2.11 STEP BY STEP PROCEDURE FOR WILCOXON TEST FOR LARGE SAMPLE

When both sample sizes are greater than about 20, the sampling distribution of U is (for practical purposes) normal.

As with the Mann Whitney U test, the sampling distribution of the statistics (In this case *T*) approaches the normal distribution as the sample size becomes large. Therefore, under these conditions, again one can perform a z-test as follows:

$Z= \{T- (N(N+1) / 4)\}/ N(N+1)(2N+1) / \sqrt{24}$

The significance decisions are identical to those for the Mann Whitney largesample case. Thus, if we have a two tailed test, the observed z is significant at the 5 per cent level if it exceeds 1.96. For the one-tailed test, significance is attained if z exceeds 1.64. However, as with the Mann-Whitney test, and for the same reasons, the Wilcoxon test tends to be restricted to use with relatively small samples.

Self Assessment Questions

What unit of sample is considered as an appropriate sample for Mann Whitney U test for small sample?Give the underlying assumptions of Mann Whitney U test?

.....

.....

.....

.....

2.12 COMPUTING THE WILCOXON SIGNED RANK SUM TEST IN SPSS

- Choose Analyse
- Select Non-parametric Tests
- Select 2 Related Samples
- Specify which two variables comprise your pairs of observation by clicking on them both then clicking on the arrow to put them under Test Pair(s) List.
- Under Test Type select Wilcoxon

If you want exact probabilities (i.e. based on the binomial distribution), click on Exact, choose Exact, then Continue

Click on OK

2.13 COMPARISON OF MANN-WHITNEY 'U' TEST AND WILCOXON MPSR TEST WITH T-TEST

The power efficiency of the Mann-Whitney and Wilcoxon tests, whilst usually somewhat lower than the corresponding *t*-test, compares very favourably with it. The Mann-Whitney and Wilcoxon tests can be used in situations where the *t* –Test would be inappropriate (e.g. where the assumptions of the *t*-test obviously do not apply). In other words, they are capable of wider application.

Different statisticians give different advice as to the relative merits of parametric and non-parametric tests. The non-parametric camp claims that their tests are simpler to compute, have fewer assumptions and can be used more widely. The parametric camp claims that their tests are robust with respect to violations of their assumptions and have greater power efficiency.

The strategy recommended here is to use the *t*-test unless the data is in form of ranks, or where the sample is small and either the distribution is obviously non-normal or there are obviously large differences in variance.

However, if you are particularly pressed for time or have a large number of analyses to do there is particularly nothing inappropriate about using non-parametric statistics, even in cases where *t*-tests might have been used.

2.14 LET US SUM UP

Two Sample test can be of two types independent sample test (two different samples being tested on one variable wherein one sample does not affect the other sample) or paired or dependent sample (same sample being tested twice or sample have some relation with each other).

The *t*-test is the parametric test for a two sample test, in non-parametric tests, Mann-Whitney 'U' test and Wilcoxon test are used for independent and paired sample respectively.

Both these tests have their own advantages, and can be used for a smaller sample size, do not have too many assumptions and can be used more widely.

2.15 UNIT END QUESTIONS

1) A researcher had an experimental group of m = 3 cases and a control group of n = 4 cases. The scores were as following:

   Exprimental scores: 9, 11, 15

   Control scores: 6, 8, 10, 13

2) In the problem 1 above, Assume these groups are independent, apply appropriate statistics and state whether the experimental condition and control conditions differ or not.

3) In the problem 1 above, Assume these groups are correlated, apply appropriate statistics and state whether the experimental condition and control conditions differ or not.

4) Doctor Radical, a math instructor at Logarithm University, has two classes in advanced calculus. There are six students in Class 1 and seven students in Class 2.

The instructor uses a programmed textbook in Class 1 and a conventional textbook in Class 2. At the end of the semester, in order to determine if the type of text employed influences student performance, Dr. Radical has another math instructor, Dr. Root, to rank the 13 students in the two classes with respect to math ability. The rankings of the students in the two classes follow:

Class 1: 1, 3, 5, 7, 11, 13

Class 2: 2, 4, 6, 8, 9, 10, 12

(Assume the lower the rank the better the student).

5) To 4 above, Apply appropriate statistics and tell if the type of text employed influenced students performance?

6) Why should you not use the large-sample *z*-test version of a non-parametric test when you have samples small enough to allow the use of small sample version?

7) Identify the non-parametric test that ought to be used.

8) You have 5 independent groups of subjects, with different numbers per group. There is also substantial departure from homogeneity of variance. The null hypothesis states that there are no differences between the groups.

You have the same situation described in question 4 (a); and in addition, the alternative hypothesis states that when the mean ranks for the 5 groups are listed from smallest to largest, they will appear in a particular *pre-specified* order.

2.16 SUGGESTED READINGS

Daniel, W. W. (1990) *Applied Non-parametric Statistics*, 2d ed. Boston: PWS-Kent.

Johnson, Morrell, and Schick (1992), Two-Sample Non-parametric Estimation and Confidence Intervals Under Truncation, *Biometrics*, 48, 1043-1056.

Siegel S. and Castellan N.J. (1988) *Non-parametric Statistics for the Behavioral Sciences* (2nd edition). New York: McGraw Hill.

Wampold BE & Drew CJ. (1990) *Theory and Application of Statistics*. New York: McGraw-Hill.

UNIT 3 KRUSKAL WALLIS ANALYSIS OF VARIANCE

Structure

3.0 Introduction

3.1 Objectives

3.2 Analysis of Variance

3.3 Introduction to Kruskal Wallis ANOVA Test

3.4 Relevant Background Information on Kruskal Wallis ANOVA Test

3.5 Step by Step Procedure for Kruskal Wallis ANOVA

3.6 Considerations for Large Sample

3.7 Comparison of ANOVA and Kruskal Wallis ANOVA Test

3.8 Let Us Sum Up

3.9 Unit End Questions

3.10 Suggested Readings and References

3.0 INTRODUCTION

So far in Unit 2 we have studied appropriate statistical tests when we wish to compare two groups (t test if data is from a normal population, Mann-Whitney U test or Wilcoxon test if there are no assumptions about the distribution of the data), but what if there are more than two groups that require comparison? One may think that we may apply the same tests in that condition too. Like for example, if there are three groups say A, B, and C, one may see the difference between A&B, B&C and A&C. This may not look so cumbersome. Now, think if we need to compare 5 groups, A, B, C, D, E, the number for comparison tests we need to do would be 10 (A&B, A&C, A&D, A&E, B&C, B&D, B&E, C&D, C&E, D&E). And what if we need to compare 6 groups? Number of two sample test in these cases become too cumbersome and may not be feasible at all. This may further lead to unnecessary calculations and also give rise to type I error. The answer in these cases when we have more than two groups (>2 groups) to be compared is to conduct Analysis of Variance.

3.0 OBJECTIVES

After reading this unit, you will be able to:

- Define ANOVA tests;
- Describe the procedure for ANOVA calculations;
- Explain Kruskal Wallis ANOVA;
- Enumerate the conditions when this test can be applied; and
- Analyse Kruskal Wallis Anova with one way ANOVA of parametric test.

3.2 ANALYSIS OF VARIANCE

The term analysis of variance (for which the acronym ANOVA is often employed) describes a group of inferential statistical procedures developed by the British statistician Sir Ronald Fisher. Analysis of variance is all about examining the amount of variability in a *y* (response) variable and trying to understand where that variability is coming from. One way that you can use ANOVA is to compare several populations regarding some quantitative variable, *y*. The populations you want to compare constitute different groups (denoted by an *x* variable), such as political affiliations, age groups, or different brands of a product. ANOVA is also particularly suitable for situations involving an experiment where you apply certain treatments *(x)* to subjects, and you measure a response *(y)*.

Null hypothesis H_O%: Population means are equal. There will be no difference in the population means.

ì1 = ì2 = ì3 = ì4

Alternative hypothesis: *h*1

Population means are not equal. There will be difference in the means of the different populations.

The logic used in ANOVA to compare means of multiple groups is similar to that used with the t-test to compare means of two independent groups. When one way ANOVA is applied to the special case of two groups, this one way ANOVA gives identical results as the t-test.

Not surprisingly, the assumptions needed for the t-test are also needed for ANOVA. We need to assume:

1) random, independent sampling from the k populations;

2) normal population distributions;

3) equal variances within the k populations.

Assumption 1 is crucial for any inferential statistic. As with the t-test, Assumptions 2 and 3 can be relaxed when large samples are used, and Assumption 3 can be relaxed when the sample sizes are roughly the same for each group even for small samples. (If there are extreme outliers or errors in the data, we need to deal with them first.)

Self Assessment Questions

1) Fill in the blanks

   i) We would use _____, if we are testing a hypothesis of ì1 = ì2 and _____Test when ì1 = ì2= ì3 = ì4 if the populations under consideration are normally distributed.

   ii) ANOVA was developed by British Statistician _____.

   iii) ANOVA is used when k _____.

   iv) ANOVA compares multiple means but the logic behind ANOVA is similar to _____ test that compares two independent means.

2) What are the assumptions of ANOVA?

   .....

.....

.....

.....

3) Why are multiple *t*-tests not preferred when we have to compare more than 2 means?

.....

.....

.....

.....

3.3 INTRODUCTION TO KRUSKAL-WALLIS ANOVA TEST

When there are more than two groups or k number of groups to be compared, ANOVA is utilised but, again since ANOVA is a parametric statistics and requires assumption of normality as a key assumption, we need to also be aware of its non-parametric counterpart. The Kruskal-Wallis test compares the medians of several (more than two) populations to see whether they are all the same or not. The Kruskal Wallis test is a non-parametric analogue to ANOVA. It can be viewed as ANOVA based on rank transformed data.

That is, the initial data are transformed to their associated ranks before being subjected to ANOVA. In other words, it's like ANOVA, except that it is computed with medians and not means. It can also be viewed as a test of medians.

The null and alternative hypotheses may be stated as:

*H*0: the population medians are equal

*H*1: the population medians differ

3.4 RELEVANT BACKGROUND INFORMATION ON KRUSKAL WALLIS ANOVA TEST

The Kruskal-Wallis one way analysis of variance by ranks (Kruskal, 1952) and (Kruskal and Wallis, 1952) is employed with ordinal (rank order) data in a hypothesis testing situation involving a design with two or more independent samples. The test is an extension of the Mann-Whitney *U* test (Test 12) to a design involving more than two independent samples and, when k = 2, the Kruskal-Wallis one way analysis of variance by ranks will yield a result that is equivalent to that obtained with the Mann-Whitney *U* test.

If the result of the Kruskal-Wallis one-way analysis of variance by ranks is significant, it indicates there is a significant difference between at least two of the sample medians in the set of k medians. As a result of the latter, the researcher can conclude there is a high likelihood that at least two of the samples represent populations with different median values.

In employing the Kruskal-Wallis one-way analysis of variance by ranks one of the following is true with regard to the rank order data that are evaluated:

a) The data are in a rank-order format, since it is the only format in which scores are available; or

b) The data have been transformed into a rank-order format from an interval/ratio format, since the researcher has reason to believe that one or more of the assumptions of the single-factor between-subjects analysis of variance (which is the parametric analog of the Kruskal-Wallis test) are saliently violated.

It should be noted that when a researcher decides to transform a set of interval ratio data into ranks, information is sacrificed. This latter fact accounts for why there is reluctance among some researchers to employ non-parametric tests such as the Kruskal Wallis oneway analysis of variance by ranks, even if there is reason to believe that one or more of the assumptions of the single factor between subjects analysis of variance have been violated.

Various sources {e.g., Conover (1980, 1999), Daniel (1990), and Marascuilo and McSweeney (1977)} note that the Kruskal Wallis one way analysis of variance by ranks is based on the following assumptions:

a) Each sample has been randomly selected from the population it represents;

b) The k samples are independent of one another;

c) The dependent variable (which is subsequently ranked) is a continuous random variable. In truth, this assumption, which is common to many non-parametric tests, is often not adhered to, in that such tests are often employed with a dependent variable which represents a discrete random variable; and

d) The underlying distributions from which the samples are derived are identical in shape.

The shapes of the underlying population distributions, however, do not have to be normal.

Maxwell and Delaney (1990) point out that the assumption of identically shaped distributions implies equal dispersion of data within each distribution. Because of this, they note that, like the single factor between subjects analysis of variance, the Kruskal Wallis one way analysis of variance by ranks assumes homogeneity of variance with respect to the underlying population distribution. Because the latter assumption is not generally acknowledged for the Kruskal Wallis one way analysis of variance by ranks, it is not uncommon for sources to state that violation of the homogeneity of variance assumption justifies use of the Kruskal Wallis one way analysis of variance by ranks in lieu of the single factor between subjects analysis of variance.

It should be pointed out, however, that there is some empirical research which suggests that the sampling distribution for the Kruskal Wallis test statistic is not as affected by violation of the homogeneity of variance assumption as is the F distribution (which is the sampling distribution for the single-factor between-subjects analysis of variance).

One reason cited by various sources for employing the Kruskal Wallis one way analysis of variance by ranks is that by virtue of ranking interval/ratio data a researcher can reduce or eliminate the impact of outliers. As noted earlier in t test for two independent samples, since outliers can dramatically influence variability, they can be responsible for heterogeneity of variance between two or more samples. In addition, outliers can have a dramatic impact on the value of a sample mean.

Zimmerman and Zumbo (1993) note that the result obtained with the Kruskal-Wallis

one-way analysis of variance by ranks is equivalent (in terms ofthe derived probability value) to that which will be obtained if the rank-orders employed for the Kruskal-Wallis test are evaluated with a single-factor between-subjects analysis of variance.

Self Assessment Questions

4) Fill in the balnks:

   a) ANOVA is a parametric statistics its equivalent non-parametric statistics is _____

   b) Kruskal Wallis ANOVA was developed by _____ and _____ in 1952.

   c) ANOVA compares means of more than two groups whereas _____ of more than two groups is compared by Kruskal Wallis ANOVA.

   d) One of the assumptions in Kruskal Wallis AONA is that the dependent variable (which is subsequently ranked) is a _____

   e) Kruskal Wallis ANOVA can be viewed as ANOVA based on _____ transformed data.

5) State the null and alternative hypothesis for Kruskal Wallis ANOVA.

   .....

   .....

   .....

   .....

6) Enumerate the assumptions of Kruskal Wallis ANOVA

   .....

   .....

   .....

   .....

3.5 STEP BY STEP PROCEDURE FOR KRUSKAL WALLIS ANOVA

1) Rank all the numbers in the entire data set from smallest to largest (using all samples combined); in the case of ties, use the average of the ranks that the values would have normally been given.

2) Total the ranks for each of the samples; call those totals $T1, T2, \ldots, Tk$, where k is the number of groups or populations.

3) Calculate the Kruskal-Wallis test statistic,

   $H = [\ 12 / N\ (N+1)\ ]\ [\ \Sigma((\Sigma R)^2 / n)\ ] - 3(N + 1)$

   N = the total number of cases

   n = the number of cases in a given group

   $(\Sigma R)^2$ = the sum of the ranks squared for a given group of subjects

4) Find the *p*-value.

5) Make your conclusion about whether you can reject Ho by examining the *p*-value.

Example of a Small Sample:

In a Study, 12 participants were divided into three groups of 4 each, they were subjected to three different conditions, A (Low Noise), B(Avearge Noise), and C(Loud Noise). They were given a test and the errors committed by them on the test were noted and are given in the table below.

Participant No.	Condition A (Low Noise)	Participant No.	Condition B (Average Noise)	Participant No.	Condition C (Loud Noise)
1	3	5	2	9	10
2	5	6	7	10	8
3	6	7	9	11	7
4	3	8	8	12	11

The researcher wishes to know whether these three conditions differ amongst themselves. and there are no assumptions of the probability. To apply Kruskal Wallis test, following steps would be taken:

Step 1: Rank all the numbers in the entire data set from smallest to largest (using all samples combined); in the case of ties, use the average of the ranks that the values would have normally been given.

Condition A	Ranks T 1	Condition B	Ranks T2	Condition C	Ranks T3
3	2.5	2	1	10	11
5	4	7	6.5	8	8.5
6	5	9	10	7	6.5
3	2.5	8	8.5	11	12
	ΣT1 = 14		ΣT2 = 26		ΣT3 = 38

Step 2: Total the ranks for each of the samples; call those totals $T1, T2, \ldots, Tk$, where k is the number of populations.

T1 =14

T2 =26

T3=38

Step3: Caculate H

$H = [12 / N (N+1)] [\Sigma((\Sigma R)^2 / n)] – 3(N + 1)$

N = 12

n = 4

$(\Sigma R)^2 = (14+ 26+ 38)^2 = 6084$

$H= [12/ 12 (12 + 1)] [(14^2/4) + (26^2/4) + (38^2/4)] – 3 (12+ 1)$

$H= [12/156] [49 + 169 + 361] – 39$

$H= (0.076 \times 579) – 39$

$H= 44.525 – 39$

$H= 5.537$

Step 4: Find the *p*-value.

Since the groups are three and number of items in each group are 4, therefore looking in table H (k=3, sample size of 4,4,4) it can be seen that the critical value is 5.692 ($\alpha = 0.05$).

Step 5: Make your conclusion about whether you can reject Ho by examining the *p*-value.

Since the critical value is more than the actual value we *accept the null hypothesis* that all the three conditions A (Low Noise), B(Avearge Noise), and C(Loud Noise), do not differ from each other, therefore, in the said experiment there was no differences in the groups performance based on the noise level.

3.6 CONSIDERATIONS FOR LARGE SAMPLE

When the number of sample increases, the table H is unable to give us with the critical values, like for example it gives critical values up to 8 samples when k=3, 4 when k=4, and 3 samples when k=5, therefore as the sample increases table H is not of use for the critical value. In such a case we resort to chi square table for getting our information on the critical value taking degrees of freedom (k – 1).

Exact tables of the Kruskal-Wallis distribution: Although an exact probability value can be computed for obtaining a configuration of ranks which is equivalent to or more extreme than the configuration observed in the data evaluated with the Kruskal-Wallis one-way analysis of variance by ranks, the chi-square distribution is generally employed to estimate the latter probability. As the values of *k* and *N* increase, the chi-square distribution provides a more accurate estimate of the exact Kruskal-Wallis distribution. Although most sources employ the chi-square approximation regardless of the values of *k* and *N*, some sources recommend that exact tables be employed under certain conditions. Beyer (1968), Daniel, and Siegel and Castellan (1988) provide exact Kruskal-Wallis probabilities for whenever $k= 3$ and the number of subjects in any of the samples is five or less. Use of the chi-square distribution for small sample sizes will generally result in a slight decrease in the power of the test (i.e., there is a higher likelihood of retaining a false null hypothesis). Thus, for small sample sizes the tabled critical chi-square value should, in actuality, are a little lower than the value of Table H.

Worked Example for a large sample

A state court administrator asked the 24 court coordinators in the state's three largest counties to rate their relative need for training in case flow management on a Likert scale (1 to 7).

1 = no training need

7 = critical training need

Training Need of Court Coordinators

County A	County B	County C
3	7	4
1	6	2
3	5	5
1	7	1
5	3	6
4	1	7
4	6	
2	4	
	4	
	5	

Step 1: Rank order the total groups' Likert scores from lowest to highest.

If tied scores are encountered, sum the tied positions and divide by the number of tied scores. Assign this rank to each of the tied scores.

Scores & Ranks Across the Three Counties

Ratings	Ranks	Ratings	Ranks
1	2.5	4	12
1	2.5	4	12
1	2.5	5	16.5
1	2.5	5	16.5
2	5.5	5	16.5
2	5.5	5	16.5
3	8	6	20
3	8	6	20
3	8	6	20
4	12	7	23
4	12	7	23
4	12	7	23

Calculating the ranks of tied scores

Example: Three court administrators rated their need for training as a 3. These three scores occupy the rank positions 7, 8, & 9.

$$(7 + 8 + 9) / 3 = 8$$

Step 2 Sum the ranks for each group and square the sums

County A		County B		County C	
Rating	Rank	Rating	Rank	Rating	Rank
3	8	7	23	4	12
1	2.5	6	20	2	5.5
3	8	5	16.5	5	16.5
1	2.5	7	23	1	2.5
5	16.5	3	8	6	20
4	12	1	2.5	7	23
4	12	6	20		
2	5.5	4	12		
		4	12		
		5	16.5		
Σ R $(\Sigma R)^2$	67.0 4489		153.5 23562.25		79.5 6320.25

Step 3 Calculate H

$H = [12 / N (N+1)] [\Sigma((\Sigma R)^2 / n)] – 3(N + 1)$

$H = [12 / 24 (24+1)] [4489 / 8 + 23562.25 / 10 + 6320.25 / 6] – 3 (24 + 1)$

$H = (0.02) (3970.725) – (75)$

$H = 4.42$

$df = (k – 1) = (3 – 1) = 2$

Interpretation

The critical chi-square table value of H for $\alpha = 0.05$, and df = 2, is 5.991

Since 4.42 < 5.991, the null hypothesis is accepted. There is *no difference* in the training needs of the court coordinators in the three counties

Self Assessment Questions

1) Rearrange the following steps of Kruskal-Wallis test in appropriate order:

   i) Calculate H

   ii) Make your conclusion about whether you can reject Ho by examining the *p*-value.

   iii) Rank all the numbers in the entire data set from smallest to largest

   iv) Find the *p*-value.

   v) Total the ranks for each of the samples; call those totals *T*1, *T*2, . . ., *Tk*, where *k* is the number of populations.

2) Fill in the Blanks

i) As the values of *k* and *N* increase, the _____ distribution provides a more accurate estimate of the exact Kruskal-Wallis distribution.

ii) Use of the chi-square distribution for small sample sizes will generally result in a slight _____ in the power of the test.

iii) When the critical value of H is more than the actual obtained value of H, we _____ the null hypothesis.

iv) When the critical value of H is less than the actual obtained value of H, we _____ the null hypothesis.

3.7 COMPARISON OF ANOVA AND KRUSKAL WALLIS ANOVA TEST

The Kruskal-Wallis (KW) ANOVA is the non-parametric equivalent of a one-way ANOVA. As it does not assume normality, the KW ANOVA tests the null hypothesis of no difference between three or more group medians, against the alternative hypothesis that a significant difference exists between the medians. The KW ANOVA is basically an extension of the Wilcoxon-Mann-Whitney (WMW) 2 sample test, and so has the same assumptions: 1) the groups have the same spreads; and 2) the data distributions have the same shape.

ANOVA compares means of different population to indicate the similarity between the populations KW ANOVA compares medians of these populations. ANOVA compares the data itself, KW ANOVA, converts data into ranks and then does its computation, in this respect Kruskal Wallis ANOVA is knows as Kruskal Wallis Analysis of Variance by Ranks.

Let's look at the Example to see how their calculations differ:

Three groups 1, 2, and 3, performed a task, we want to see whether they differ or not.

Group 1	Group 2	Group 3		Group 1 (Ranks)	Group 2 (Ranks)	Group 3 (Ranks)
3	6	9		1	3.5	10
5	7	10		2	5.5	13
6	8	11		3.5	7.5	15.5
7	9	12		5.5	10	17
8	10	15		7.5	13	18
9	10			10	13	
	11				15.5	
38	61	57	Total	29.5	68	73.5

ANOVA

$n_1 = 6$ $n_2 = 7$ $n_3 = 5$

$\Sigma X_1 = 38$ $X_2 = 61$ $\Sigma X_3 = 57$

$\Sigma(X_1)^2 = 264$ $\Sigma (X_2)^2 = 551$ $\Sigma (X_3)^2 = 671$

$SS_{total} = (264 + 551 + 671) – [(38 + 61+ 57)^2 / 18] = 134$

$SS_{Between\ Groups} = (38^2/6) + (61^2/7) + (57^2/5) – [(38 + 61+ 57)^2 / 18]$

$SS_{Within\ Gropus} = [\ 264 - (38^2/6)] + [\ 551- (61^2/7)] + [\ 671- (57^2/5)] = 63.962$

Source of Variation	SS	df	MS	F ratio	F critical Value	Test Decision
Between Groups	70.038	2	35.019	8.213	3.68	Reject H_0
Within Groups	63.962	15	4.264			
Total	134.000	17				

Kruskal Wallis H test:

$H = [12 / 18(18+1)] [(29.5^2/6) + (68^2/7) + (73.5^2/5)] – [3 (18+1)]$

$H= 66.177 – 57 = 9.177$

Chi Square for Degrees of freedom 2 (3 – 1) is 5.99, Therefore reject the H_o

In both the case, ANOVA or Kruskal Wallis ANOVA, we will reject the Null Hypothesis, and state that the three groups differ.

F ratio as a function of H:

The fisher's F or F ratio or ANOVA one way variance is equivalent to H test or Kruskal Wallis test or Kruskal Wallis ANOVA, or ANOVA by rank order. This can also be seen in book Iman and Conover (1981)

Where the rank transform statistics states:

$F= [\{(k-1)/(N-k)\} \{((N-1)/H)-1\}]^{-1}$

If We see from the above mentioned example

F was 8.213 and H was 9.177

$F= [\{(3-1)/(18-3)\} \{((18-1)/9.177)-1\}]^{-1}$

$F= [(2/15) \{(17/9.177)-1\}]^{-1}$

$F= [0.133 \times (1.852-1)]^{-1} = 0.1214^{-1}$

F= 8.231

3.8 LET US SUM UP

Kruskal Wallis One way Analysis of Variance (KW ANOVA) is a Non-parametric Analogue of ANOVA one way variance for independent sample. Kruskal Wallis ANOVA is used when there are more than 2 groups (k > 2). The assumption of KW ANOVA are that: a) Each sample has been randomly selected from the population it represents; b) The k samples are independent of one another; c) The dependent variable (which is subsequently ranked) is a continuous random variable. In truth, this assumption, which is common to many non-parametric tests, is often not adhered to, in that such tests are often employed with a dependent variable which represents a discrete random variable; and d) the underlying distributions from which the samples are derived are identical in shape.

The ANOVA or F test and KW ANOVA Or H test are equivalent to each other and can be appropriately used depending upon the type of population in question.

3.9 UNIT END QUESTIONS

1) Under what circumstances does the chi-square distribution provide an appropriate characterisation of the sampling distribution of the Kruskal–Wallis *H* statistic?

2) Data were collected from three populations—*A, B,* and *C,*—by means of a completely randomized design.

   The following describes the sample data:

   nA = nB = nC = 15

   RA = 235 RB = 439 RC = 361

   a) Specify the null and alternative hypotheses that should be used in conducting a test of hypothesis to determine whether the probability distributions of populations *A, B,* and *C* differ in location.

   b) Conduct the test of part a.

3) A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

Observation	*Program 1*	*Program 2*	*Program 3*	*Program 4*
1	9	10	12	9
2	12	6	14	8
3	14	9	11	11
4	11	9	13	7
5	13	10	11	8

Calculate H, and report your results appropriately

4) An economist wants to test whether mean housing prices are the same regardless of which of 3 air-pollution levels typically prevails. A random sample of house purchases in 3 areas yields the price data below.

Mean Housing Prices (Thousands of Dollars):

MEAN HOUSING PRICES (THOUSANDS OF DOLLARS): *Pollution Level*			
Observation	Low	Mod	High
1	120	61	40
2	68	59	55
3	40	110	73
4	95	75	45
5	83	80	64

Calculate H and report your results with the p-value of 0.05

5) Show that H is equivalent to the F test statistics in one way analysis of variance problem if applied to the ranks of the observation rather than the actual numbers. (**Hint**: Express the F ratio as a function of H)

3.10 SUGGESTED READING AND REFERENCES

Daniel, W. W. (1990) *Applied Non-parametric Statistics*, 2d ed. Boston: PWS-Kent.

Iman, R. L., and W. J. Conover (1981), Rank transformations as a bridge between parametric and non-parametric statistics, The American Statistician, 35, 124–129.

Siegel S. and Castellan N.J. (1988) *Non-parametric Statistics for the Behavioral Sciences* (2nd edition). New York: McGraw Hill.

References

Johnson, Morrell, and Schick (1992), Two-Sample Non-parametric Estimation and Confidence Intervals Under Truncation, *Biometrics*, 48, 1043-1056.

Leach, C. (1979). Introduction to statistics: A non-parametric approach for the social sciences. Chichester: John Wiley & Sons

Lehman, E. L. (1975). Non-parametric statistical methods based on ranks. San Francisco: Holden-Day.

Wampold BE & Drew CJ. (1990) Theory and application of statistics. New York: McGraw-Hill.

UNIT 4 CHI-SQUARE AND KENDALL RANK CORRELATION

Structure

4.0 Introduction

4.1 Objectives

4.2 Concept of Correlation

- 4.2.1 Scatter Plot
- 4.2.2 Characteristics of Correlation

4.3 Measures of Correlation

- 4.3.1 Parametric Statistics
- 4.3.2 Non-parametric Statistics

4.4 Kendall's Rank Order Correlation (Kendall's tau): (π)

- 4.4.1 Relevant Background Information on Test

4.5 Step by Step Procedure for Kendall Rank Order Correlation

4.6 Further Considerations about Kendall's Tau

- 4.6.1 Comparison of Rho and Tau
- 4.6.2 Efficiency of Rho

4.7 Chi-Square Test

4.8 Relevant Background Information on Test

4.9 Step by Step Procedure for Chi-Square Test

4.10 Further Considerations about Chi-Square

4.11 Let Us Sum Up

4.12 Unit End Questions

4.13 Suggested Readings

4.0 INTRODUCTION

In this unit, we will be discussing about the issues relating to the association and relationship between two or more variables. Generally when we want to measure the linear relationship between two variables, we apply Product Moment Coefficient of Correlation to the data and compute the 'r' value and check for its significance. This again we would do so if the data is normally distributed and the measurement of scores etc. are atleast in interval scale and there is a large sample. However if the sample size is small, and the distribution of the data is not known and the measurement is in nominal or ordinal scale, then we use non-parametric statistics related correlation, as for example the Rho or the Kendall Tau or where we need to know the association between two variables we may use the chi square test. In this unit we will be presenting first the measures of correlation both in parametric and non-parametric statistics, followed by Kendall rank order correlation, the Spearman Rank order correlation and the Chi Square test.

4.1 OBJECTIVES

On completing this unit, you will be able to:

- Define parametric and non-parametric tests of correlation;
- Explain the concepts underlying the non-parametric correlations;
- Describe the different non-parametric correlation techniques;
- Enumerate the step by step calculation of Kendall Tau; and
- Enumerate the step by step calculation of Chi Square test.

4.2 CONCEPT OF CORRELATION

The term "correlation" refers to a process for establishing whether or not relationships exist between two variables. Correlation quantifies the extent to which two quantitative variables, X and Y, "go together." When high values of X are associated with high values of Y, a positive correlation exists. When high values of X are associated with low values of Y, a negative correlation exists. If values of X increases bringing about an increase in the values of Y simultaneously, X and Y are said to be positively correlated. If increases in X values bring about comparative decreases in the values of Y, then X and Y are said to be negatively correlated. If there is no typical trend in the increase or decrease of the variables then it is said to be not correlated or having zero correlation. Correlation ranges from -1 to 0 to +1. Correlation of +1.00 will indicate a perfect positive correlation and -1 will indicate a perfect negative correlation. Between these two extremes there could be many other degrees of correlation indicating positive or negative relationship between the variables. The correlation cannot exceed 1 in either direction. But it can have 0.54, 0.82, or 0.24, or 0.63 and so on at the positive level and at the negative level , it can have -0.55, -0.98, -0.67, -0.27 etc. All the latter are negative correlations and will not go beyond -1.00. Similarly the correlations that were mentioned as positive , will not exceed +1.00.

4.2.1 Scatter Plot

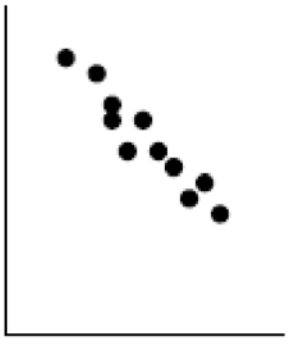
The first step is creating a scatter plot of the data. "There is no excuse for failing to plot and look."

In general, scatter plots may reveal a

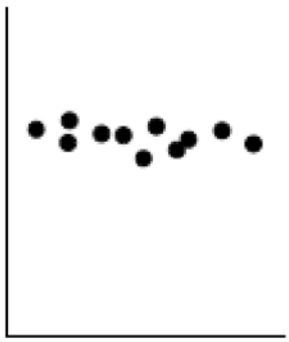
- positive correlation (high values of X associated with high values of Y)
- negative correlation (high values of X associated with low values of Y)
- no correlation (values of X are not at all predictive of values of Y).

These patterns are demonstrated in the figure below

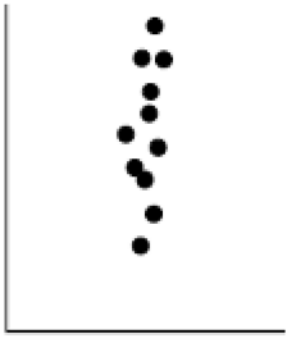
(A) Positive Correlation



(B) Negative Correlation



(A) No Correlation



(B) No Correlation

Correlation Coefficient

A single summary number that gives you a good idea about how closely one variable is related to another variable

This summary answers the following questions:

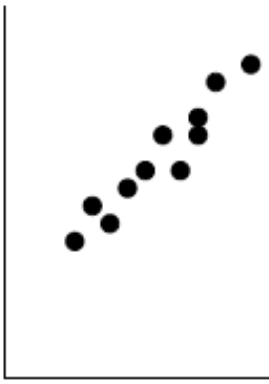
a) Does a relationship exist?

b) If so, is it a positive or a negative relationship? and

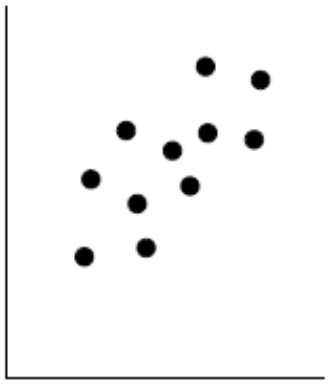
c) Is it a strong or a weak relationship?

Additionally, the same summary number would allow us to make accurate predictions about one variable when we have knowledge about the other variable.

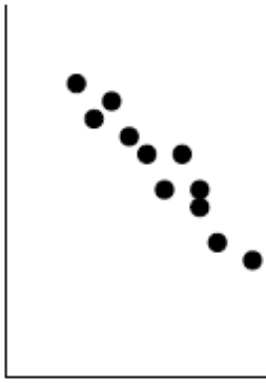
Correlation coefficients (denoted by *r*) are statistics that quantify the relationship between X and Y in unit free terms. When all points of a scatter plot fall directly on a line with an upward incline, $r = +1.00$, but when all points fall directly on a downward incline, $r =$



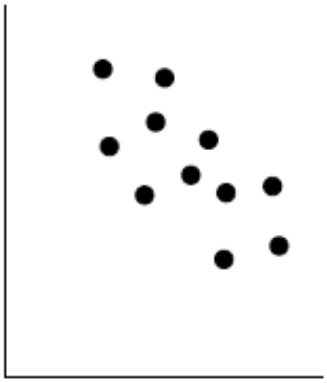
(A) Strong Positive Correlation



(B) Weak Positive Correlation



(C) Strong Negative Correlation

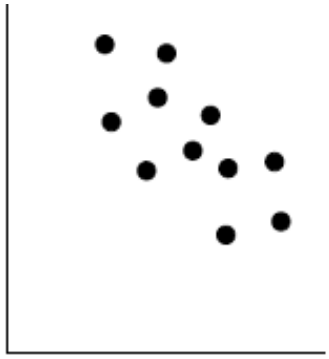


(D) Weak Negative Correlation

(A) Strong Positive Correlation

(B) Weak Positive Correlation

(A) Strong Negative Correlation



(B) Weak Negative Correlation

It is seen from the above that the strong correlation at both positive and negative directions is almost in a line with all the dots are placed very close to each other. On the other hand, the weak positive or negative correlation (refer to the graph above on the right hand side) that the points are placed far away from each other though the direction is somewhat clear. Thus there is a correlation but it appears rather weak.

4.2.2 Characteristics of Correlation

1) They tell you the direction of the relationship between two variables.

   If your correlation coefficient is a negative number you can tell, just by looking at it, that there is a negative relationship between the two variables. As you may recall from the last chapter, a negative relationship means that as values on one variable increases (go up) the values on the other variable tend to decrease (go down) in a predictable manner.

   If your correlation coefficient is a positive number, then you know that you have a positive relationship. This means that as one variable increases (or decreases) the values of the other variable tend to go in the same direction. If one increases, so does the other. If one decreases, so does the other in a predictable manner.

2) Correlation Coefficients always fall Between -1.00 and +1.00

   All correlation coefficients range from -1.00 to +1.00. A correlation coefficient of -1.00 tells you that there is a *perfect negative relationship* between the two variables. This means that as values on one variable *increase* there is a perfectly predictable *decrease* in values on the other variable. In other words, as one variable goes up, the other goes in the opposite direction (it goes down).

   A correlation coefficient of +1.00 tells you that there is a *perfect positive relationship* between the two variables. This means that as values on one variable *increase* there is a perfectly predictable *increase* in values on the other variable. In other words, as one variable goes up, so does the other.

   A correlation coefficient of 0.00 tells you that there is a zero correlation, or no relationship, between the two variables. In other words, as one variable changes (goes up or down) you can't really say anything about what happens to the other variable. Sometimes the other variable goes up and sometimes it goes down. However, these changes are not predictable.

3) Larger Correlation Coefficients Mean Stronger Relationships

   Most correlation coefficients (assuming there really is a relationship between the two variables you are examining) tend to be somewhat lower than plus or minus 1.00 (meaning that they are not perfect relationships) but are somewhat above 0.00. Remember that a correlation coefficient of 0.00 means that there is no relationship between the two variables based on the data given .

The closer a correlation coefficient is to 0.00, the weaker is the relationship and the less able one is to tell exactly what happens to one variable based on the knowledge of the other variable. The closer a correlation coefficient approaches plus or minus 1.00 the stronger the relationship is and the more accurately you are able to predict what happens to one variable based on the knowledge you have of the other variable.

4.3 MEASURES OF CORRELATION

4.3.1 Parametric Statistics

a) Pearson product moment correlation coefficient (Most widely accepted as a single appropriate statistics for correlation)

4.3.2 Non-parametric Statistics

a) Spearman's rank order correlation coefficient: Better known as "Spearman Rho" (Siegel & Castellan, 1988) assumes that the variables under consideration were measured on at least an ordinal (rank order) scale, that is, that the individual observations can be ranked into two ordered series. Spearman R can be thought of as the regular Pearson product moment correlation coefficient, that is, in terms of proportion of variability accounted for, except that Spearman R is computed from ranks.

b) Kendall's Tau: Explained in section 4.3.

c) Chi Square (Categorical Variables): Explained in section 4.7

Self Assessment Questions

1) Fill in the blanks:

   i) Scatter plots may reveal a _____correlation (high values of X associated with high values of Y)

   ii) Scatter plots may reveal a _____correlation (high values of X associated with low values of Y)

   iii) Scatter plots may reveal _____correlation (values of X are not at all predictive of values of Y).

   iv) Correlation coefficients range from _____ to _____

   v) A correlation coefficient of _____ tells you that there is a *perfect positive relationship* between the two variables.

   vi) The closer a correlation coefficient is to 0.00, the _____ the relationship

   vii) Correlation coefficient is a single summary number that gives you a good idea about *how closely one variable is* _____ *to another variable*

2) What questions does correlation coefficient answers?

   .....

   .....

   .....

   .....

3) Name any two methods for calculating correlation?

.....

.....

.....

.....

4.4 KENDALL'S RANK ORDER CORRELATION (KENDALL'S TAU): (ð)

Kendall's tau (**ð)** is one of a number of measures of correlation or association. Measures of correlation are not inferential statistical tests, but are, instead, descriptive statistical measures which represent the degree of relationship between two or more variables. Upon computing a measure of correlation, it is a common practice to employ one or more inferential statistical tests in order to evaluate one or more hypotheses concerning the correlation coefficient. The hypothesis stated below is the most commonly evaluated hypothesis for Kendall's tau.

Null Hypothesis

H_o: **ð** = *0*

(In the underlying population the sample represents, the correlation between the ranks of subjects on Variable X and Variable Y equals 0.)

4.4.1 Relevant Background Information on Test

Prior to reading the material in this section the reader should review the general discussion of correlation, of the Pearson product moment correlation coefficient and Spearman's rank order correlation coefficient (which also evaluates whether a monotonic relationship exists between two sets of ranks). Developed by Kendall (1938), tau is a bivariate measure of correlation/association that is employed with rank-order data. The population parameter estimated by the correlation coefficient will be represented by the notation **ð** (which is the lower case Greek letter tau). As is the case with Spearman's rank-order correlation coefficient Rho (***ñ***), Kendall's tau can be employed to evaluate data in which a researcher has scores for n subjects/objects on two variables (designated as the X and Y variables), both of which are rank-ordered.

Kendall's tau is also commonly employed to evaluate the degree of agreement between the rankings of m = 2 judges for n subjects/objects. As is the case with Spearman's rho, the range of possible values Kendall's tau can assume is defined by the limits - 1 to +1 (i.e., - 1 < r > +1). Although Kendall's tau and Spearman's rho share certain properties in common with one another, they employ a different logic with respect to how they evaluate the degree of association between two variables.

Kendall's tau measures the degree of agreement between two sets of ranks with respect to the relative ordering of all possible pairs of subject/objects.

One set of ranks represents the ranks on the X variable, and the other set represents the ranks on the Y variable.

Specifically, The data are in the form of the following two pairs of observations expressed in a rank-order format:

a) (R_x, R_y,) (which, respectively, represent the ranks on Variables X and Y for the 1st subject/object); and

b) (R_{xj}, R_{yj},) (which, respectively, represent the ranks on Variables X and Y for the jth subject/object).

If the sign/direction of the difference ($R x_i - R_{yj}$), that is a pair of ranks is said to be concordant (i.e. in agreement).

If the sign/direction of the difference ($R_{xi} - R_{xj}$), a pair of ranks is said to be discordant (i.e., disagree).

If ($R_{yi} - R_{yj}$) and/or ($R_{xi} - R_{xj}$) result in the value of zero, a pair of ranks is neither the concordant nor discordant.

Kendall's tau is a proportion which represents the difference between the proportions of concordant pairs of ranks less the proportion of discordant pairs of ranks.

The computed value of tau will equal + 1 when there is complete agreement among the rankings (i.e., all of the pairs of ranks are concordant), and will equal -1 when there is complete disagreement among the rankings (i.e., all of the pairs of ranks are discordant).

As a result of the different logic involved in computing Kendall's tau and Spearman's rho, the two measures have different underlying scales, and, because of this, it is not possible to determine the exact value of one measure if the value of the other measure is known.

In spite of the differences between Kendall's tau and Spearman's rho, the two statistics employ the same amount of information, and, because of this, it is equally likely to detect a significant effect in a population.

In contrast to Kendall's tau, Spearman's rho is more commonly discussed in statistics books as a bivariate measure of correlation for ranked data. Two reasons for this are as follows:

a) The computations required for computing tau are more tedious than those required for computing rho; and

b) When a sample is derived from a bivariate normal distribution.

4.5 STEP BY STEP PROCEDURE FOR KENDALL RANK-ORDER CORRELATION

These are the steps in use of the Kendall rank order Correlation coefficient **ð(tau)**:

Rank the observations on the X variable from 1 to N. Rank the observations on the Y variable from 1 to N.

Arrange the list of N subjects so that the rank of the subjects on variable X are in their natural order, that is, 1, 2, 3,….N.

Observe the Y ranks in the order in which they occur when X ranks are in natural order. Determine the value of S, the number of agreements in order minus the number of disagreements in order, for the observed order of the Y ranks.

If there are no ties among either the X or the Y observations then we use the formula:

T = 2S / (N (N -1))

Where:

S = (score of agreement – score of disagreement on X and Y)

N = Number of objects or individuals ranked on both X and Y

If there are ties then the formula would be:

$T = 2S / [N (N-1) - T_x$ $\quad T = 2S / [N (N-1) - Ty$

Where:

S and N are as above

$T_x = \Sigma t (t - 1)$, t being the number of tied observations in each group of the ties on the X variable

$T_y = \Sigma t (t - 1)$, t being the number of tied observation in each group of the ties on the Y variable

If the N subjects constitute a random sample from some population, one may test the hypothesis that the variable X and Y are independent in that population. The method for doing so depends on the size of N:

For $N \leq 10$, Table — Upper tail probabilities for T, the Kendall rank order correlation coefficient

For $N > 10$, but less than 30, Table – Critical value for T, the Kendall rank order correlation coefficient

For $N < 30$ (or for intermediate significance levels for $10 < N \leq 30$) compute the value of z associated with T by using formula given below and use the z table

$z = 3T \quad N (N - 1) / 2 (2N+5)$

If the probability yielded by the appropriate method is equal to or less than the critical value, null hypothesis may be rejected in the favour of alternative hypothesis.

Worked up Example:

Without Ties:

Suppose we ask X and Y to rate their preference for four objects and give points out of 10. Now to see whether their preferences are related to each other we may use the following steps:

Data:

	A	B	C	D
X	6	8	5	2
Y	8	4	9	6

Step 1: Ranking the data of X and Y

	A	B	C	D
X	3	4	2	1
Y	3	1	4	2

Step 2: Rearrange the data of X in order of 1 to N (4 in this case)

	D	C	A	B
X	1	2	3	4

Step 3: Put the corresponding score of Y in order of X and Determine number of agreements and disagreements

	D	C	A	B
X	1	2	3	4
Y	2	4	3	1

To calculate S we need number of agreements and disagreements. This can be calculated by

Using the Y scores, starting from left and counting the number of ranks to its right that are larger, these are agreements in order. We subtract from this the number of ranks to its right that are smaller- these are the disagreements in order. If we do this for all the ranks and then sum the results we obtain S:

Y	2	4	3	1	Total
	2	+	+	-	+1
		4	-	-	-2
			3	-	-1
				1	0
				Grand Total= S	- 2

Step 4: Calculate T

$T = 2S / (N (N -1))$

$T = 2 (– 2) / (4 (4 – 1))$

$T = – 4 / 12$

$T = – 0.33$

Thus, $T = – 0.33$ is a measure of the agreement between the preferences of X and Y.

With Ties:

The two set of ranks to be correlated are:

Subject	A	B	C	D	E	F	G	H	I	J	K	L
Status striving rank	3	4	2	1	8	11	10	6	7	12	5	9
Yielding rank	1.5	1.5	3.5	3.5	5	6	7	8	9	10.5	10.5	12

As usual we would first rearrange X and observe the scores of corresponding Y scores to calculate S

Subject	D	C	A	B	K	H	I	E	L	G	F	J	
Status striving rank	1	2	3	4	5	6	7	8	9	10	11	12	
Yielding rank	3.5	3.5	1.5	1.5	10.5	8	9	5	12	7	6	10.5	Total
	3.5	0	-	-	+	+	+	+	+	+	+	+	8
		3.5	-	-	+	+	+	+	+	+	+	+	8
			1.5	0	+	+	+	+	+	+	+	+	8
				1.5	+	+	+	+	+	+	+	+	8
					10.5	-	-	-	+	-	-	0	-4
						8	+	-	+	-	-	+	0
							9	-	+	-	-	+	-1
								5	+	+	+	+	4
									12	-	-	-	-3
										7	-	+	0
											6	+	1
												10.5	0
											S=	Grand Total	25

We compute the value of S in usual way

S = (8-2) + (8-2) + (8-0) + (8-0) + (1-5) +

(3-3) + (2-3) + (4-0) + (0-3) + (1-1) + (1-0) = 25

It should be noted that, when there are tied observations, the ranks will be tied and neither rank in comparison pair precedes the other, so a value of 0 is assigned in the computation of S.

Having determined that S = 25, we now determine the value of T_x and T_Y. There are no ties among the scores on social status striving, i.e. in the X ranks and thus $T_x = 0$

On Y scores there are three sets of tied ranks. Two subjects are tied at 1.5, two subjects at 3.5, and two subjects' at 10.5 ranks. In each of these cases T = 2, the number of tied observations. Thus may be computed as:

$T_Y = \Sigma\, t\,(t-1)$

$=2\,(2-1) + 2(2-1) + 2(2-1)$

$= 6$

With $T_x = 0$, $T_Y = 6$, S = 25, and N =12, we may determine the value of T by using formula:

$$T = 2S / [\sqrt{N\,(N-1) - T_x}\ \sqrt{N\,(N-1) - T_y}\ ]$$

$$T = (2 \times 25) / \sqrt{12(12-1) - 0}\ \sqrt{12(12-1) - 6}$$

$= 0.39$

If we had not corrected the above coefficient for ties, i.e. we had used the previous formula for computing T we would have found T = 0.38. Observe that the effect of correcting for ties is relatively small unless the proportion of tied ranks is large or the number of ties in a group of ties is large.

4.6 FURTHER CONSIDERATIONS ABOUT KENDALL'S TAU

4.6.1 Comparison of Rho and Tau

For the example of tied observation if one calculates r it will be 0.62, whereas the T is 0.39. This example illustrates the fact that T and r have different underlying scales, and numerically they are not directly comparable to each other. That is if we measure the degree of correlation between A and B by using r and then do the same for A and C by using T, we cannot then say whether A is more closely related to B or to C because we have used noncomparable measures of correlation. It should be noted, however that there is a relation between the two measures which is best expressed in the following inequality:

$$-1 \leq 3\,T - 2r \leq 1$$

There are also differences in interpretation of the two measures. The spearman rank order correlation coefficient rho ***(ñ)*** is the same as a Pearson product moment correlation coefficient computed between variables the values of which consists of ranks. On the other hand, the Kendall rank-order correlation coefficient(**ð=tau)** has a different interpretation. It is the difference between the probability that, in the observed data, X and Y are in the same order and the probability that the X and Y data are in different orders. T_{XY} is different in the relative frequencies in the sample.

However, both coefficients utilise the same amount of information in the data, and thus both have the same sensitivity to detect the existence of association in the population. That is, the sampling distributions of T and r are such that for a given set of data both will lead to rejection of the null hypothesis at the same level of significance. However it should be remembered that the measures are different and measure association in different ways.

4.6.2 Efficiency of Rho

The Spearman Rho ***(ñ)*** and The Kendall (tau=ð) are similar in their ability to reject H_o, inasmuch as they make similar use of the information in the data.

When used on data to which the Pearson product moment correlation coefficient r is properly applicable, both Rho (***(ñ)*** and tau (**ð)** have efficiency of 91 percent. That is, Rho is approximately as sensitive a test of independence of two variables in a bivariate normal population with a sample of 100 cases as the Pearson r with 91 cases (Moran,1).

Self Assessment Questions

1) Fill in the blanks:

   i) Rho and tau have different underlying scales, and numerically they are not _____ to each other.

   ii) Developed by _____ in year _____, tau is a _____measure of correlation/association that is employed with rank-order data.

2) State true or false:

   i) Kendall's tau measures the degree of agreement between two sets of ranks with respect to the relative ordering of all possible pairs of subject/objects.

   ii) Kendall's tau and Spearman's rho, the two measures have different underlying scales, and, because of this, it is not possible to determine the exact value of one measure if the value of the other measure is known.

   iii) Kendall's tau and Pearson's r both are rank order correlation, therefore both can be compared.

4.7 CHI-SQUARE TEST

The chi-square (X^2) test measures the alignment between two sets of frequency measures. These must be categorical counts and *not* percentages *or* ratios measures (for these, use another correlation test).

Note that the frequency numbers should be significant and be at least above 5 (although an occasional lower figure may be possible, as long as they are not a part of a pattern of low figures).

Chi Square performs two types of functions:

1) **Goodness of fit**

A common use is to assess whether a measured/observed set of measures follows an expected pattern. The expected frequency may be determined from prior knowledge (such as a previous year's exam results) or by calculation of an average from the given data.

The null hypothesis, H_0 is that the two sets of measures are not significantly different.

2) **Measure of Independence**

The chi-square test can be used in the reverse manner to goodness of fit. If the two sets of measures are compared, then just as you can show they align, you can also determine if they do *not* align.

The null hypothesis here is that the two sets of measures are similar.

The main difference in goodness-of-fit vs. independence assessments is in the use of the Chi Square table. For goodness of fit, attention is on 0.05, 0.01 or 0.001 figures. For independence, it is on 0.95 or 0.99 figures (this is why the table has two ends to it).

4.8 RELEVANT BACKGROUND INFORMATION ON TEST

The chi-square goodness-of-fit test, also referred to as the chi-square test for a single sample, is employed in a hypothesis testing situation involving a single sample. Based on some pre existing characteristic or measure of performance, each of *n* observations (subjects/objects) that is randomly selected from a population consisting of N observations (subjects/objects) is assigned to one of k mutually exclusive categories.' The data are summarized in the form of a table consisting of k cells, each cell representing one of the k categories.

The experimental hypothesis evaluated with the chi-square goodness-of-fit test is whether or not there is a difference between the observed frequencies of the k cells and their expected frequencies (also referred to as the theoretical frequencies). The expected frequency of a cell is determined through the use of probability theory or is based on some pre existing empirical information about the variable under study. If the result of the chi-square goodness-of-fit test is significant, the researcher can conclude that in the underlying population represented by the sample there is a high likelihood that the observed frequency for at least one of the k cells is not equal to the expected frequency of the cell. It should be noted that, in actuality, the test statistic for the chi-square goodness-of-fit test provides an approximation of a binomially distributed variable (when k = 2) and a multinomially distributed variable (when k > 2). The larger the value of *n*, the more accurate the chi-square approximation of the binomial and multinomial distributions.

The chi-square goodness-of-fit test is based on the following assumptions:

a) Categorical nominal data are employed in the analysis. This assumption reflects the fact that the test data should represent frequencies for k mutually exclusive categories;

b) The data that are evaluated consists of a random sample of n independent observations. This assumption reflects the fact that each observation can only be represented once in the data; and

c) The expected frequency of each cell is 5 or greater.

When this assumption is violated, it is recommended that if $k = 2$, the binomial sign test for a single sample be employed to evaluate the data. When the expected frequency of one or more cells is less than 5 and $k > 2$, the multinomial distribution should be employed to evaluate the data. The reader should be aware of the fact that sources are not in agreement with respect to the minimum acceptable value for an expected frequency.

Many sources employ criteria suggested by Cochran (1952), who stated that none of the expected frequencies should be less than 1 and that no more than 20% of the expected frequencies should be less than 5. However, many sources suggest the latter criteria may be overly conservative. In the event that a researcher believes that one or more expected cell frequencies are too small, two or more cells can be combined with one another to increase the values of the expected frequencies.

Zar (1999) provides an interesting discussion on the issue of the lowest acceptable value for an expected frequency. Within the framework of his discussion, Zar (1999) cites studies indicating that when the chi-square goodness-of-fit test is employed to evaluate a hypothesis regarding a uniform distribution, the test is extremely robust.

A robust test is one that still provides reliable information, in spite of the fact that one or more of its assumptions have been violated. A uniform distribution (also referred to as a rectangular distribution) is one in which each of the possible values a variable can assume has an equal likelihood of occurring. In the case of an analysis involving the chi-square goodness-of-fit test, a distribution is uniform if each of the cells has the same expected frequency.

4.9 STEP BY STEP PROCEDURE FOR CHI-SQUARE TEST

1) Write the observed frequencies in column *O*

2) Figure the expected frequencies and write them in column *E*.

Expected Frequencies:

When you find the value for chi square, you determine whether the observed frequencies differ significantly from the expected frequencies. You find the expected frequencies for chi square in three ways:

1) You hypothesize that all the frequencies are equal in each category. For example, you might expect that half of the entering freshmen class of 200 at Tech College will be identified as women and half as men. You figure the expected frequency by dividing the number in the sample by the number of categories. In this exam pie, where there are 200 entering freshmen and two categories, male and female, you divide your sample of 200 by 2, the number of categories, to get 100 (expected frequencies) in each category.

2) You determine the expected frequencies on the basis of some prior knowledge. Let us use the Tech College example again, but this time pretend we have prior knowledge of the frequencies of men and women in each category from last year's entering class, when 60% of the freshmen were men and 40% were women. This year you might expect that 60% of the total would be men and 40% would be women. You find the expected frequencies by multiplying the sample size by each of the hypothesized population proportions. If the freshmen total were 200, you would expect 120 to be men (60% x 200) and 80 to be women (40% x 200).

3) Use the formula to find the chi-square value:

   Chi Square = $\Sigma\ [(O - E)^2 / E]$

   Where:

   O is the Observed Frequency in each category

   E is the Expected Frequency in the corresponding category

4) Find the *df*. (*N*-1)

5) Find the table value (consult the Chi Square Table.)

6) If your chi-square value is *equal to or greater than* the table value, reject the null hypothesis: *differences in your data are not due to chance alone*

Worked Up Example:

Situation: Mr. X., the manager of a car dealership, did not want to stock cars that were bought less frequently because of their unpopular color. The five colors that he ordered were red, yellow, green, blue, and white. According to Mr. X, the expected frequencies or number of customers choosing each color should follow the percentages of last year. She felt 20% would choose yellow, 30% would choose red, 10% would choose green, 10% would choose blue, and 30% would choose white. She now took a random sample of 150 customers and asked them their colour preferences. The results of this poll are shown in Table below under the column labelled as observed frequencies."

Category Color	Observed Frequencies	Expected Frequencies
Yellow	35	30
Red	50	45
Green	30	15
Blue	10	15
White	25	45

The expected frequencies in Table are figured from last year's percentages. Based on the percentages for last year, we would expect 20% to choose yellow. Figure the expected frequencies for yellow by taking 20% of the 150 customers, getting an expected frequency of 30 people for this category. For the colour red we would expect 30% out of 150 or 45 people to fall in this category.

Using this method, Thai figured out the expected frequencies 30, 45, *15*, *15*, and 45. Obviously, there are discrepancies between the colours preferred by customers in the poll taken by Mr.X. and the colours preferred by the customers who bought their cars *last* year. Most striking is the difference in the green and white colours. If Thai were to follow the results of her poll, she would stock twice as many green cars than if she were to follow the customer colour preference for green based on last year's sales. In the case of white cars, she would stock half as many this year. What to do? Mr. X. needs to know whether or not the discrepancies between last year's choices (expected frequencies) and this year's preferences on the basis of his poll (observed frequencies) demonstrate a *real* change in customer colour preferences. It could be that the differences are simply a result of the random sample she *chanced to* select. If so, then the population of customers really has not changed from last year as far as colour preferences go.

The *null hypothesis* states that there is no significant difference between the expected and observed frequencies.

The *alternative hypothesis* states they *are* different. The level of significance (the point at which you can say with 95% confidence that the difference is NOT due to chance alone) is set at .05 (the standard for most science experiments.) The chi-square formula used on these data is

Chi Square = $\Sigma\ [(O - E)^2 / E]$

Where:

O is the Observed Frequency in each category

E is the Expected Frequency in the corresponding category

df is the "degree of freedom" (n-1)

We are now ready to use our formula for X^2 and find out if there *is* a significant difference

between the observed and expected frequencies for the customers in choosing cars. We will set up a worksheet; then you will follow the directions to form the columns and solve the formula.

1) *Directions for Setting up Worksheet for Chi Square*

Category	O	E	O-E	$(O-E)^2$	$(O-E)^2$ / E
Yellow	35	30	5	25	0.83
Red	50	45	5	25	0.56
Green	30	15	15	225	15
Blue	10	15	-5	25	1.67
White	25	45	-20	400	8.89
				Total=	26.95

This Total is the Chi Square value. After calculating the Chi Square value, find the *"Degrees of Freedom."*

(Remember: DO *NOT* SQUARE THE NUMBER YOU GET, NOR FIND THE SQUARE ROOT - THE NUMBER YOU GET FROM COMPLETING THE CALCULATIONS AS ABOVE IS CHI SQUARE.)

2) *Degrees of freedom (df)* refers to the number of values that are free to vary after restriction has been placed on the data. For instance, if you have four numbers with the restriction that their sum has to be 50, then three of these numbers can be anything, they are free to vary, but the fourth number *definitely* is restricted. For example, the first three numbers could be 15, 20, and 5, adding up to 40; then the fourth number has to be 10 in order that they sum to 50. The degrees of freedom for these values are then three. The degrees of freedom here is defined as N - 1, the number in the group minus one restriction (4 - 1).

3) Find the table value for Chi Square. Begin by finding the *df* found in step 2 along the left hand side of the table. Run your fingers across the proper row until you reach the predetermined level of significance (.05) at the column heading on the top of the table. The table value for Chi Square in the correct box of *4 df* and *P=.05* level of significance is 9.49.

4) If the calculated chi-square value for the set of data you are analysing (26.95) is equal to or greater than the table value (9.49), reject the null hypothesis. *There is a significant difference between the data sets that cannot be due to chance alone.* If the number you calculate is LESS than the number you find on the table, then you can probably say that any differences are due to chance alone.

In this situation, the rejection of the null hypothesis means that the differences between the expected frequencies (based upon last year's car sales) and the observed frequencies (based upon this year's poll taken by Mr.X) are not due to chance. That is, they are not due to chance variation in the sample Mr.X took. There is a real difference between them. Therefore, in deciding what colour autos to stock, it would be to Mr.X's advantage to pay careful attention to the results of her poll!

Another Example:

Let us take an example of Males and Females in two different categories, full stop and rolling stop and no stop. Now to see whether they are different from each other or more similar to each other we will follow the following steps

Step 1: Add numbers across columns and rows. Calculate total number in chart.

Unobtrusive Male Versus Female

	Male	Female	
Full Stop	6	6	= 12
Rolling Stop	16	15	= 31
No Stop	4	3	= 7
	= 26	= 24	= 50

Step 2: Calculate the expected numbers for each individual cell. Do this by multiplying row sum by column sum and dividing by total number. For example: using 1st cell in table (Male/Full Stop);

12 x 26 / 50 = 6.24

2nd cell in table (Female/Full Stop):

12 x 24 / 50 = 5.76

Step 3: Now you should have an observed number and expected number for each cell. The observed number is the number already in 1st chart. The expected number is the number found in the last step (step 2). Sometimes writing both numbers in the chart can be helpful

	Male	Female	
Full Stop	6 (observed) 6.24 (expected)	6 (observed) 5.76 (expected)	= 12
Rolling Stop	16 (observed) 16.12 (expected)	15 (observed) 14.88 (expected)	= 31
No Stop	4 (observed) 3.64 (expected)	3 (observed) 3.36 (expected)	= 7
	= 26	= 24	= 50

Step 4:

Chi Square = Sum of (Observed - Expected)^2 / Expected

Calculate this formula for each cell, one at a time. For example, cell #1 (Male/Full Stop):

Observed number is: 6 Expected number is: 6.24

Plugging this into the formula, you have:

$(6 – 6.24)^2 / 6.24 = .0092$

Continue doing this for the rest of the cells, and add the final numbers for each cell together for the final Chi Square number. There are 6 total cells, so at the end you should be adding six numbers together for you final Chi Square number.

Step 5: Calculate degrees of freedom (*df*):

(Number of Rows – 1) x (Number of Columns – 1)

(3 – 1) x (2 – 1)

2 x 1 =

2 *df* (degrees of freedom)

Step 6: Look up the number in the chart at end of handout. At .05 significance level, with 2 *df*, the number in chart should be 5.99. Therefore, in order to reject the null hypothesis, the final answer to the Chi Square must be greater or equal to 5.99. The Chi Square/final answer found was .0952. This number is less than 5.99, so you fail to reject the null hypothesis, thus there is no difference in these groups.

4.10 FURTHER CONSIDERATIONS ABOUT CHI SQUARE

Observations must appear in one cell only. For instance, if we looked at male and female swimmers and hurdlers, one person could appear in both the swimmers *and* the hurdlers category if they enjoyed both sports. This would make use of Chi square invalid. Actual frequencies must appear in the cells, not percentages, proportions or numbers which do anything other than count. For instance, the mean of an interval scale variable cannot appear.

LOW expected frequencies

One limitation is that one should not proceed with a chi square test where expected frequency cells fall below 5. The rule of thumb which most statisticians inherited, and which comes from Cochran (1954) which was that *no more than 20% of expected cells should fall below* 5. This would rule out any 2 X 2 in which at least one expected cell was less than 5.

Hypothetical table:

Age	Conversed	Did not converse	Total
5 years	2	6	8
7 years	6	2	8
Total	8	8	16

For total sample sizes less than 20 and two expected cells below 5, the risk of a type I error is too high. For instance, the data shown in hypothetical table above give a *chi square* of 4.0 (which is 'significant' for one *df)* yet it's easy to see, again, without much formal statistical training, that the result was relatively likely to occur - only two children in each age group needed to move away, in opposite directions, from the expected frequencies of four in each cell for these results to occur. From first principles (working out all the possible combinations) the probability of these results occurring comes out substantially higher than 0.05. If you have these sort of data it doesn't take too long to work from first principles but it's far better to make sure your analysis will be valid by taking a large enough sample, with a sensible design. Even with tables larger than 2X2, if several expected frequencies fall below 5 and the row or column total are quite severely skewed, the possibility of a type I error increases.

Self Assessment Questions

1) What are the assumptions of chi-square goodness-of-fit test?

.....

.....

.....

.....

2) Chi square performs two major functions, what are these?

.....

.....

.....

.....

3) State true or false:

i) The expected frequency of a cell is determined through the use of probability theory or is based on some pre existing empirical information about the variable under study.

ii) If several expected frequencies fall below 5, the possibility of a type II error increases.

iii) The chi-square (c^2) test measures the alignment between two sets of frequency measures.

iv) "The data that are evaluated consists of a random sample of n independent observations." Is not a cardinal assumptions of chi square?

4.11 LET US SUM UP

In this unit we learnt about the concept of correlation and how parametric test is used to compute the product moment coefficient of correlation. We thedn learnt about the non-parametric tests for corrleation and leatrnt about the Rho and Tau. The Rho was by Spearman and was known as Spearman Rank Correlation while Kendall's Tau was known as ð (tau). We also learnt about how to calculate Kendall's tau and learnt about the importance of Chi-Square test. We also learnt as to how to calculate chi-square.

4.12 UNIT END QUESTIONS

1) Compute correlation coefficient for each of the following pairs of sample observations:

a)

x	33	61	20	19	40
y	26	36	65	25	35

b)

x	89	102	120	137	41
y	81	94	75	52	136

c)

x	2	15	4	10	
y	11	2	15	21	

d)

x	5	20	15	10	3
y	80	83	91	82	87

2) Compare T and r in terms of correlation and state your views?

3) Should a chi-Square test be carried out on the following data?

7 1

2 7

4) A (fictitious) Survey shows that. in a sample of 100.9 I people are against the privatisation of health services, whereas 9 support the idea.

   a) What test of significance can be performed on this data?

   b) Calculate the chi square value and check it for significance.

   c) Could this test be one-tailed?

   If for a large sample, we knew *on/y* that 87% of people were against the idea and were for could we carry out the same test to see whether this split is significant

5) What is the difference between chi square goodness of fit test and measure of independence test?

6) What do you understand by efficiency of T?

4.13 SUGGESTED READINGS

Daniel, W. W. (1990) *Applied Non-parametric Statistics*, 2d ed. Boston: PWS-Kent.

Johnson, Morrell, and Schick (1992), Two-Sample Non-parametric Estimation and Confidence Intervals Under Truncation, *Biometrics*, 48, 1043-1056.

Siegel S. and Castellan N.J. (1988) *Non-parametric Statistics for the Behavioral Sciences* (2nd edition). New York: McGraw Hill.

Wampold BE & Drew CJ. (1990) *Theory and Application of Statistics*. New York: McGraw-Hill.